

Rugg, Graham (2016) Carbon-halogen bond formation in enzymes. PhD thesis.

<http://theses.gla.ac.uk/7045/>

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Carbon-Halogen Bond Formation in Enzymes

Graham Rugg
BSc.

Submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

School of Chemistry
College of Science and Engineering
University of Glasgow

Abstract

SyrB2, a non-haem Fe halogenase first discovered in 2005, carries out a cryptic chlorination during the biosynthesis of syringomycin E in the bacterium *Pseudomonas syringae*. SyrB2 chlorinates its native substrate, L-Threonine (Thr), at an unactivated methyl group. It is able to activate this highly unreactive position using an oxoferryl intermediate of its active site complex, which abstracts a hydrogen from the substrate methyl group to form a bioradical. Whilst a provisional mechanism was quickly derived from the mechanisms of similar non-haem Fe enzymes, two features of this mechanism remain unclear: firstly, the structure or structures of the oxoferryl intermediate of its active-site complex, and secondly, why SyrB2 does not hydroxylate Thr in what would appear to be a plausible side-reaction. This latter problem is believed to be the result of substrate placement, as in reaction with two non-native substrates, α -aminobutyrate (Aba) and norvaline (Nva), SyrB2 is able to function as a hydroxylase.

This thesis sets out to answer these two questions, as well as to pursue several preliminary goals. Firstly, a method validation study was carried out on several oxoferryl model complexes, which showed that B3LYP reproduced several parameters from CASPT2 benchmarks from the literature better than other tested functionals. Next, protein-substrate interactions were studied through docking and molecular dynamics simulations, which uncovered a new position for Thr. Finally, the mechanism of SyrB2 in reaction with these three substrates was investigated in a QM/MM study, which identified two likely structures of the oxoferryl active-site complex, as well as a new species in which the substrate radical intermediate coordinates to the iron complex.

Contents

1	Introduction	11
1.1	Background	11
1.1.1	Halogenase Enzymes	11
1.2	SyrB2	15
2	Theoretical Background	19
2.1	Overview	19
2.2	Quantum Chemical Methods	19
2.2.1	Overview and the Hartree-Fock Method	19
2.2.2	Coupled Cluster Theory	22
2.2.3	Density Functional Theory	24
2.3	Molecular Mechanics	26
2.4	Hybrid QM/MM calculations	28
2.5	Geometry Optimisation	29
2.6	Molecular Dynamics	30
2.7	Docking	32
3	Electronic Structure of SyrB2's Oxoferryl Intermediate	34
3.1	Introduction	34
3.2	Computational Details	35
3.3	Results and Discussion	36
3.3.1	Calculation of Structural Parameters	36
3.3.2	Electronic Structure of Oxoferryl Complexes	37
3.3.3	Comparison of DFT to CCSD(T)	40
3.3.4	Spin-State Energetics of Other Models	44
3.4	Conclusion	46
4	The Protein-Substrate Complex	47
4.1	Introduction	47

4.1.1	The Need for a Model of the Protein-Substrate Complex	47
4.1.2	Insights into Active-Site Structure	47
4.2	Computational Details	49
4.2.1	Docking	49
4.2.2	Molecular Dynamics with Apoprotein or Fe(II) Active Site Complex	50
4.2.3	Molecular Dynamics with Fe(IV) Active-Site Complex	51
4.2.4	Analysis	51
4.3	Results and discussion	53
4.3.1	Docking	53
4.3.2	Molecular Dynamics with Iron (II) Active-Site Complex	58
4.3.3	Molecular Dynamics with Iron (IV) Active-Site Complex	63
4.3.4	Hydrogen Bonding Environment of the Oxoferryl Species	66
4.4	Conclusions	68
5	Mechanistic Calculations	70
5.1	Introduction	70
5.1.1	Oxygen Activation	71
5.1.2	Hydrogen Abstraction	73
5.2	Computational Details	74
5.3	Results and Discussion	76
5.3.1	Oxygen Activation	76
5.3.2	Hydrogen Abstraction	89
5.3.3	Radical Rebound	94
5.3.4	Summary	97
5.4	Conclusions	102
6	Conclusions	103
A	Supplementary Data from QM/MM Calculations	106
A.1	Supplementary data at the def2-TZVP Basis Level	106
A.2	Geometries at the def2-SVP Basis Level	107
A.3	Energy Profiles at the def2-SVP Basis Level	110

List of Figures

1.1	Formation of ^-OX electrophile in haem- and vanadate-dependent haloperoxidases.	12
1.2	Formation of HOX in flavin-dependent halogenases	13
1.3	Reactions of the nucleophilic halogenases	13
1.4	Reaction mechanism of SyrB2.	15
1.5	T₁ and T₂	16
1.6	Reaction of Thr, Nva and ABA with SyrB2	17
1.7	Possible isomers of the oxoferryl reactive intermediate	18
3.1	$[\text{Fe(IV)(O)(TMC)(NCCH}_3)]^{2+}$ and $[\text{Fe(IV)(O)(TMC-py)}]^{2+}$	37
3.2	Spin-state ordering of Model 1a by method	38
3.3	Model 1a in the quintet, triplet and septet spin states	39
3.4	Valence orbitals of Model 1a , shown with quintet multiplicity.	40
3.5	RHF/CCSD(T) energies of model 1b by basis set level	42
3.6	Spin-state ordering (relative to quintet) of Model 1b from CCSD(T) by method	42
3.7	Models 1b , 2b and 3b	44
4.1	Protein-substrate interactions	48
4.2	Most favourable docked conformation of THR	54
4.3	Four most favourable docked conformations of THR	55
4.4	Most favourable docked conformations of ABA and NVA	57
4.5	Hydrogen bonding partners of the THR γ -OH group.	60
4.6	Position of ABA and THR following MD equilibration	60
4.7	2OG hydrogen bonding network	61
4.8	Distribution of the $\text{C} - \text{C}^\alpha - \text{C}^\beta - \text{C}^\gamma$ torsion of Phe196	61
4.9	Phe196 in holoprotein and protein with bound ABA	62
4.10	T₁ and T₂	63
4.11	Distances d_{OH} and d_{CCl} between THR and the active-site complex isomer D1	64
4.12	Position of THR relative to the active-site complex isomers D2 and D5	65

4.13	ABA with D3 and D4 active-site complex isomers	65
4.14	Mechanisms of Fe(IV) isomerisation proposed in literature	69
5.1	Oxygen activation, hydrogen abstraction, and radical rebound	70
5.2	Proposed mechanisms of oxygen activation in NHFe enzymes	72
5.3	Exchange enhancement on the quintet surface	74
5.4	QM and inner regions of THR snapshot	75
5.5	Electronic structures of the oxygen adduct C	79
5.6	Oxygen adduct (C)	80
5.7	Energy profile of oxygen activation on the quintet surface	81
5.8	Energy profile of oxygen activation for the three substrates on the quintet surface	82
5.9	Species J	82
5.10	Comparison of energy profiles of oxygen activation for the substrate ABA .	83
5.11	Comparison of energy profiles of oxygen activation for the substrate THR .	83
5.12	Oxygen activation intermediates	84
5.13	Products of oxygen activation on the triplet surface	84
5.14	Oxoferryl species ⁵ D1 formed from oxygen activation on the quintet surface.	85
5.15	Oxygen activation in the presence of THR.	85
5.16	THR in Conformations 1 and 2	89
5.17	TS_{DE} for THR in Conformation 1	90
5.18	Rotational conformations of the ferric hydroxide group after hydrogen ab- straction from ABA	91
5.19	TS_{DE} for NVA C ^γ and C ^δ	92
5.20	Energy profile of hydrogen abstraction from the three substrates	93
5.21	Energy profile of chloride rebound	94
5.22	Species E2	95
5.23	Energy profile of chloride rebound for NVA C ^δ and THR Configuration 1, with reverse scans	96
5.24	TS_{EF} (left) and TS_{EF'} (right) for the substrate ABA	96
5.25	Energy profile of hydroxide rebound	97
5.26	Energy profile of hydrogen abstraction and radical rebound for ABA, THR Conf. 1 and NVA C ^δ	98
5.27	Energy profile of hydrogen abstraction and radical rebound for THR Confs. 1 and 2 and NVA C ^γ and C ^δ	99
A.1	Energy profile of oxygen activation for ABA at the def2-TZVP basis level .	107

A.2	Energy profile of hydrogen abstraction and radical rebound for ABA, THR Conf. 1 and NVA C^δ at the def2-SVP basis level	110
A.3	Energy profile of hydrogen abstraction and radical rebound for THR Conf. 1 and 2 and NVA C^γ and C^δ at the def2-SVP basis level	111

List of Tables

2.1	Terms included in the Autodock VINA scoring function	33
3.1	Structural parameters of $[\text{Fe(IV)(O)(TMC)(NCCH}_3)]^{2+}$	36
3.2	Structural parameters of $[\text{Fe(IV)(O)(TMC-py)}]^{2+}$	37
3.3	Spin-state ordering of Model 1a by density functional	38
3.4	Comparison of Models 1a and 1b	41
3.5	Relative energies of spin-states of Model 1b from CCSD(T)	41
3.6	Coupled-cluster diagnostics	43
3.7	Geometric parameters of model complexes	45
3.8	Spin-state orderings of Models 1b , 2b and 3b	46
4.1	Protonation states of histidine residues	49
4.2	Equilibration procedure for MD simulations	51
4.3	Docked poses of THR	56
4.4	The ten docked poses of ABA and NVA with highest affinity	56
4.5	Backbone RMSD from crystal structure	58
4.6	Protein–substrate interactions during MD simulations	58
4.7	RMSFs of flexible residues	59
4.8	Throughputs of channels T₁ and T₂	62
4.9	Placement of THR relative to oxoferryl complexes	64
4.10	Placement of ABA relative to oxoferryl complexes	66
4.11	Placement of NVA relative to oxoferryl complexes	66
4.12	Hydrogen-bonding environments of oxoferryl species	67
5.1	Comparison of model complexes to experimentally validated interatomic distances	77
5.2	Spin populations and relative spin-state energies of the oxygen adduct C .	78
5.3	Geometric parameters of the oxygen adduct C	79
5.4	Geometric parameters of the species formed during oxygen activation . . .	87

5.5	Comparison of d_{OH} and d_{CCl} between MD simulations and QM/MM calculations	88
5.6	Spin-state separations and spin populations of the oxoferryl complexes D .	88
5.7	Geometric parameters of species B to F' on the quintet surface. For species E to F' for ABA and THR Conf. 2, the parameters listed are for the lower energy rotamer of the ferric OH group. Where multiple minor conformers are possible (D for THR Confs. 1 and 2 and E for ABA), values listed are for the conformation from which the scan commenced	101
A.1	Geometric parameters of various species at the def2-TZVP level	106
A.2	Geometric parameters of the oxygen adduct C at the def2-SVP basis level	107
A.3	Geometric parameters of species B to F' on the quintet surface at the def2-SVP basis level. For ABA and THR Conf. 2 species E–F' , the parameters listed are for the lower energy rotamer of the ferric OH group. Where multiple minor conformations are possible (D for THR Confs. 1 and 2 and E for ABA), values listed are for the conformation from which the scan commenced	108
A.4	Spin-state separations and spin populations of the oxoferryl complexes D at the def2-SVP basis level.	109

Abbreviations

Abbreviation	
Thr	Threonine
Aba	α -amino-butyrate
Nva	Nor-valine
THR	Pantetheinylated threonine
ABA	Pantetheinylated α -amino-butyrate
NVA	Pantetheinylated nor-valine
PPant	Phosphopantetheine
NHFe	Non-haem Fe
FT-NHFe	Facial triad-coordinated non-haem Fe
d_{OH}	Distance from oxoferryl oxygen to nearest abstractable hydrogen at given substrate site
d_{CCl}	Distance from fe-bound chloride to given substrate carbon
2OG	2-oxo-glutarate
T_1	The T_1 diagnostic
D_1	The D_1 diagnostic
T₁	Substrate channel of SyrB2
T₂	Allosteric channel leading to SyrB2's active site
TACN	1,4,7-Triazacyclononane
MSK	Merz-Singh-Kollman
TMC	1,4,8,11-tetramethyl-1,4,8,11-tetraazacyclotetradecane
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation

Acknowledgements

I would like to take this opportunity to thank my supervisor, Dr Hans Martin Senn, for his guidance, patience, and support, and for his ability and readiness to explain complicated concepts in simple terms.

I would also like to thank my family for their support throughout the last four years. Their encouragement and positivity has made this endeavour an infinitely more pleasant experience.

Thanks are also due to all of the members of the WestCHEM computational chemistry consortium. Our weekly meetings provided a fantastic opportunity to acquire new skills and to discuss new findings in an informal setting.

Finally, I would like to thank the Engineering and Physical Sciences Research Council for my scholarship and funding.

Chapter 1

Introduction

1.1 Background

SyrB2, discovered in 2005 in the bacterium *Pseudomonas syringae*, is a non-haem Fe 2-oxoglutarate dependent halogenase. Employing a radical rebound reaction, it chlorinates an unactivated methyl carbon of L-threonine during the biosynthesis of syringomycin E. This direct chlorination of an alkane is an impressive synthetic feat. However, despite extensive study, both experimental and computational, the reaction itself is not fully understood. This thesis aims to provide a complete picture of this enzyme’s mechanism, through study of both the complicated electronic structure of the iron centre and the intermolecular interactions between the protein and substrate.

1.1.1 Halogenase Enzymes

Nature makes use of a plethora of halogenated molecules, ranging from simple halomethanes to more complicated molecules such as peptides and polyketides [1, 2]. The array of organisms that produce halogenated compounds is similarly diverse, ranging from bacteria to plants to higher animals. The majority of halogenated natural compounds contain chlorine, followed by bromine, but iodine- and fluorine- containing products are not unknown. Broadly speaking, the enzymes that create carbon-halogen bonds can be grouped into three families—electrophilic halogenases, nucleophilic halogenases, and radical halogenases.

Electrophilic Halogenases

The first family of halogenases to be discovered employed an electrophilic mechanism. The family can be subdivided into haem-dependent haloperoxidases, known since 1966 [3], vanadate-dependent haloperoxidases, known since 1983 [4, 5], and flavin dependent halogenases, first identified in 1995 [6] (although even a rudimentary understanding of

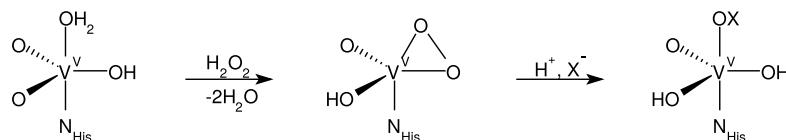
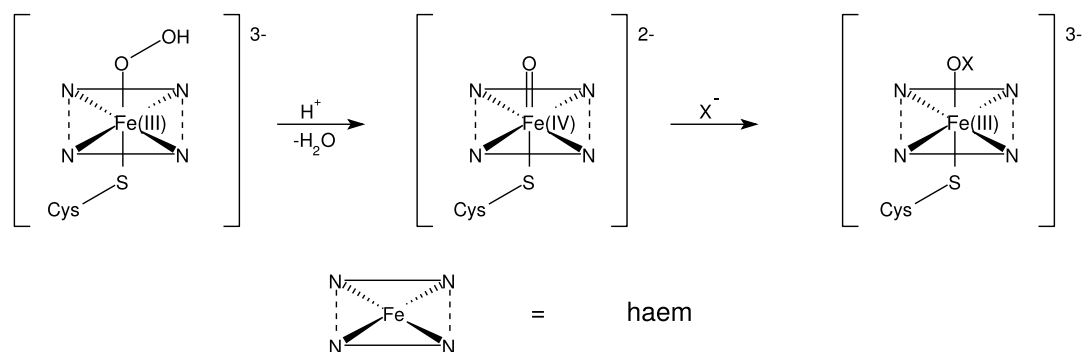


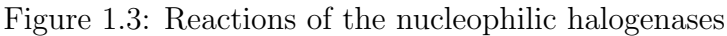
Figure 1.1: Formation of ^-OX electrophile in haem-dependent (above) and vanadate-dependent (below) haloperoxidases.

the mechanism of the flavin-dependent halogenases took several more years to gather, for a comprehensive review see [7]). Haem- and vanadate- dependent haloperoxidases are further dependent on H_2O_2 , both using the oxidation of X^- to create the electrophile XO^- (Figure 1.1), which can then attack nucleophilic substrates.

Flavin dependent halogenases rely on flavin adenine dinucleotide (FAD) as a cofactor. They employ a multi-step reaction (see Figure 1.2), first generating FADHOOH , which is in turn used to generate an electropositive halogen HOX . There has been some debate as to whether free HOX reacts directly with the substrate, or whether an N-haloamine intermediate is formed first, the latter being the currently favoured mechanism. They are extremely regio-, stereo- and substrate selective, but in spite of a wealth of experimental data, their mechanism is not well understood.

Nucleophilic Halogenases

To date, all known enzymes that facilitate nucleophilic attack by X^- have the same substrate, S-adenosyl-L-methionine (SAM). SAM fluorinase (F1A) and chlorinase (Sa1L) attack the same position of SAM, yielding adenosine derivatives (Figure 1.3). Halide methyl transferases attack the methyl group of the sulfonium centre, yielding halo-methanes. Although halide methyl transferases were discovered as early as 1998 [8], and are amongst the most common halogenating enzymes, the first crystal structure only appeared in 2010 [9] and little or no computational work has subsequently appeared [2].



Radical Halogenases

The radical halogenases use a non-haem Fe (NHFe) centre to generate a substrate radical intermediate that can react with an Fe-bound X^\bullet radical equivalent. The first representative of the family to be characterised was CmaB in 2005 [10]. In the same year, SyrB2 was first characterised [11], with a set of crystal structures appearing the following year [12]. A number of similar enzymes were subsequently discovered, including CytC3 [13], Hal [14], BarB1 and BarB2 [15], Thr3 [16], and HctB [17]. Crystal structures of CytC3 and Hal were resolved in 2009 [18] and 2010 [14] respectively. The NHFe halogenases are dependent on 2-oxoglutarate (2OG) as a co-substrate, and almost all require the substrate to be presented on a phosphopantetheinyl arm attached to a carrier protein, though a recently discovered family [19] have been shown to be able to chlorinate a free-standing substrate. Similar halogenating activity has been detected in inorganic complexes [20].

NHFe centres appear in a number of different enzymes, facilitating reactions as diverse as epoxidation, ring closure and hydroxylation (for a detailed review see [21]). SyrB2 alone, in addition to chlorination, has been shown to be capable of bromination [22], hydroxylation [23], and even nitrogenation and azidation [24]. NHFe hydroxylases have been studied rather more extensively than their halogenase cousins, with an early derivation of their mechanism proposed in 1982 [25]. The close structural and mechanistic similarities between NHFe hydroxylases and halogenases has allowed the hydroxylases to provide something of a mechanistic “template” for their halogenase cousins. A recently proposed mechanism for SyrB2 [26] is shown in Figure 1.4.

The resting state (**A**) of the active-site iron complex contains Fe(II) bound to two histidine residues, a chloride ion, 2OG and a water molecule. The binding of the substrate in the active site triggers [27] the loss of water (**B**). After oxygen binds (**C**), it attacks and decarboxylates 2OG, yielding succinate and a high-spin Fe(IV)-oxo intermediate (**D**). Next, a hydrogen atom is abstracted from the substrate, forming an Fe(III)-OH species and a substrate radical intermediate (**E**), which is chlorinated in a radical rebound reaction (**F**).

To date, due to the relative paucity of crystal structures of NHFe halogenases, the bulk of the computational study of these enzymes has been carried out on SyrB2. However, in spite of the large number of studies of this enzyme, a number of significant mechanistic questions remain unanswered.

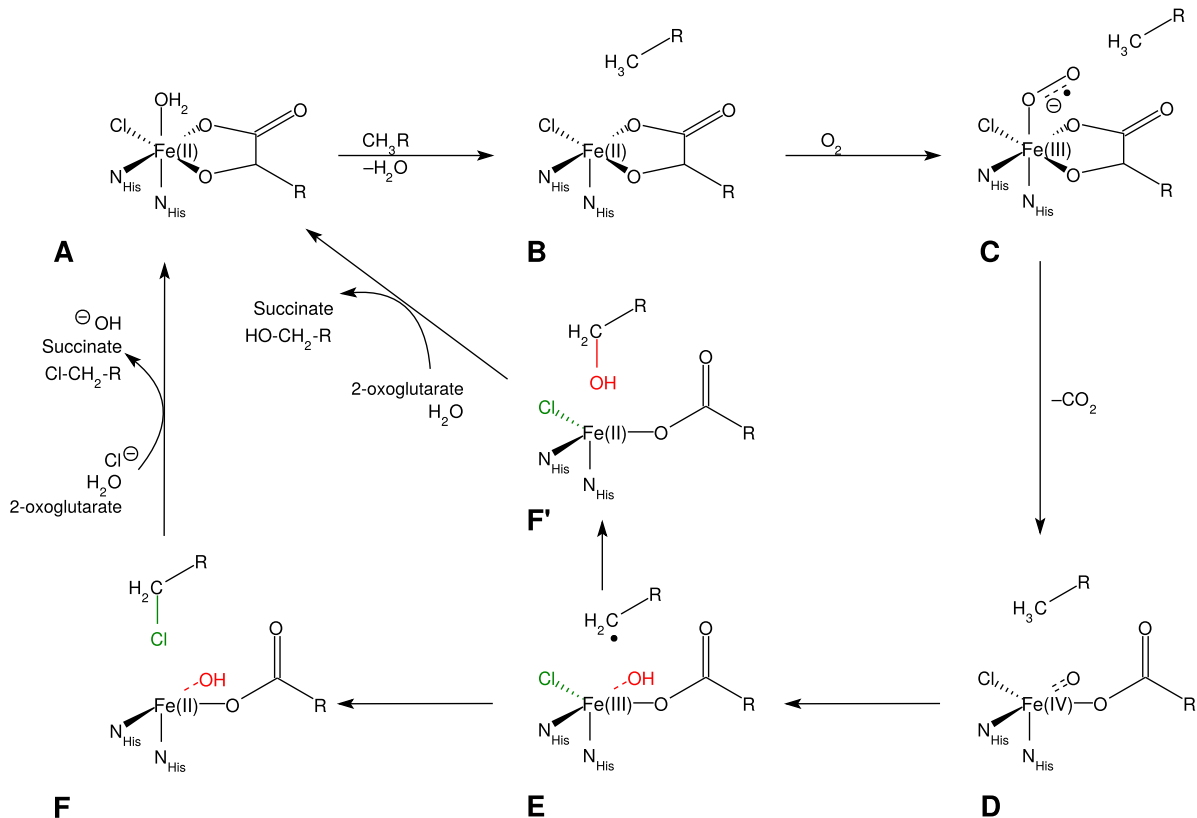


Figure 1.4: Reaction mechanism of SyrB2.

1.2 SyrB2

A wealth of experimental information about SyrB2 has been derived in the ten years since its discovery. Three crystal structures have been resolved: one shows the enzyme with bound iron, 2OG and chloride, a second shows the enzyme with bound iron, 2OG and bromide, and a third shows the enzyme with bound 2OG only. The iron complex sits in a cavity in the centre of the protein, to which two channels (**T₁** and **T₂**, Figure 1.5) were identified in the crystal structure [12]. **T₁** was regarded as the most likely to receive the substrate due to its length and breadth. A subsequent mutational analysis [16] found that mutating several residues in and around **T₁** to alanine reduced or stopped altogether the halogenating activity of the enzyme, strengthening the argument that this is the substrate channel.

There are two main features of SyrB2's reactivity that remain unexplained. The first of these is its chemoselectivity. In reaction with its native substrate, L-threonine (Thr), SyrB2 is absolutely chemoselective, yielding only 4-chloro-Thr (see Figure 1.4, step **E–F**). This is in spite of the hydroxide ligand to the iron complex at **E**, which would appear to be

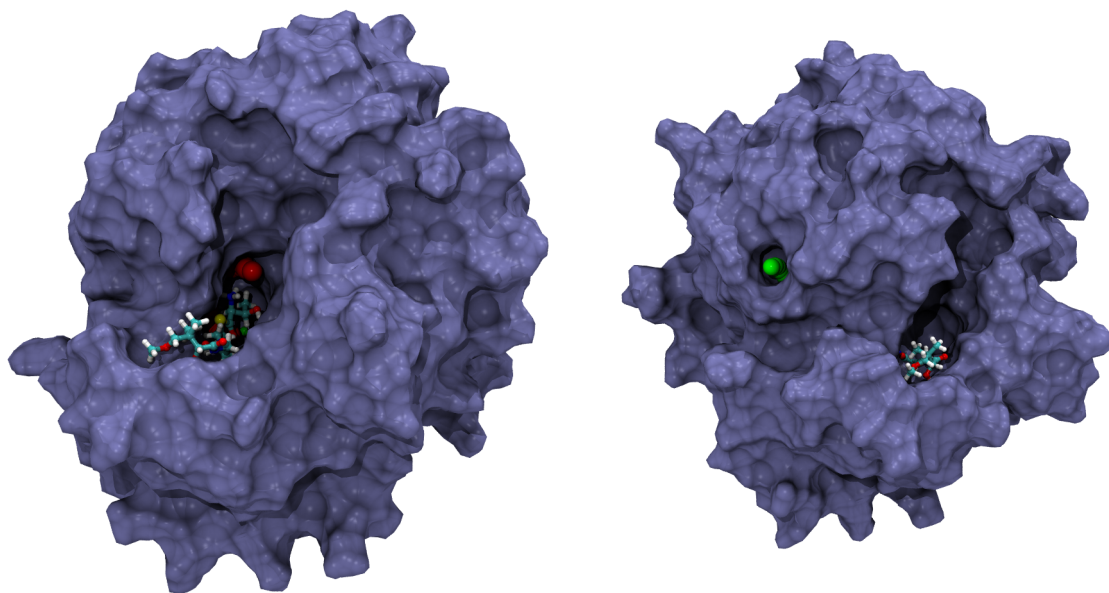


Figure 1.5: Left: **T₁**, shown with docked PPant-Thr, tunnel probe shown in red. Right: **T₂**, tunnel probe shown in green

available for radical rebound. In reaction with non-native substrates norvaline (Nva) and α -amino butyrate (Aba), a mixture of halogenated and hydroxylated products is produced (see Figure 1.6). Aba differs from Thr only in its lack of Thr's hydroxide group, Nva also lacking this hydroxide and being longer by one methylene unit, which suggests that a difference in substrate placement leads to this difference in selectivity. This is bolstered by the fact that the A118D and A118E variants of SyrB2, which have mutations that inhibit chloride binding, are still unable to function as hydroxylases to Thr [12]. The difference in substrate positions was confirmed by a recent study [28], which measured the placement of the three substrates Thr, Nva and Aba relative to a nitrosyl analogue to the species **C**. Using an NO probe, Martinie *et al.* were able to form an iron-nitrosyl adduct sufficiently stable to allow measurement of several important structural parameters through hyperfine sublevel correlation spectroscopy. These included the distance from the central iron to the hydrogen atoms of the substrates' terminal methyl groups, and the Fe–N–H angles for these hydrogens. The Fe–H distances correlated to the order of the rates of reaction, with Nva, which has the highest rate of reaction, having a lower Fe–H distance than Aba, which in turn had a lower Fe–H distance than Thr, the substrate with the lowest rate of reaction.

The other major unexplained feature of SyrB2's reaction cycle is the structure of the oxoferryl intermediate (**D**). Several isomers of this complex, with mono- or bidentate succinate and different orientations of the ligands, have been proposed. Initially, it was

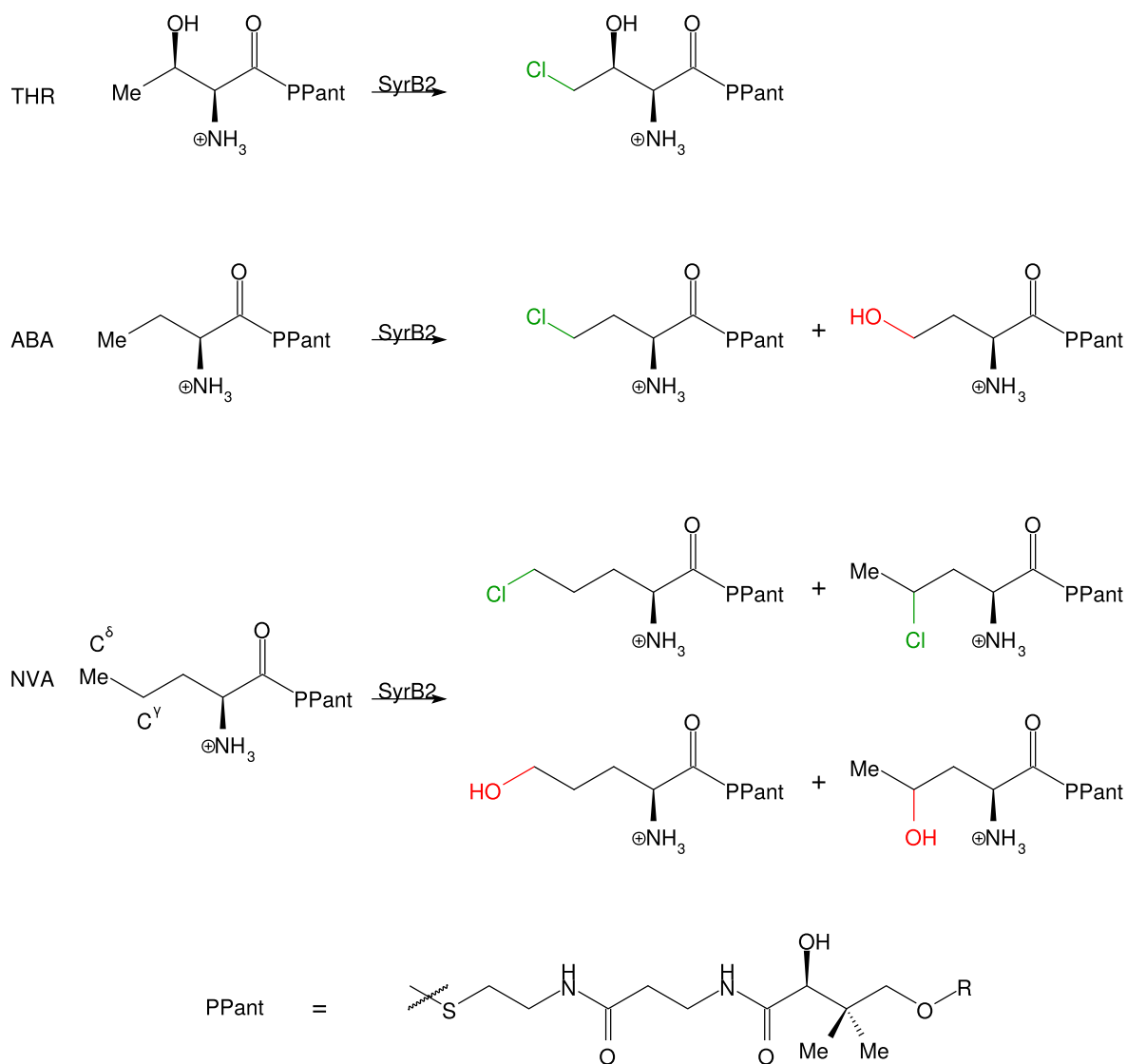


Figure 1.6: Reaction of Thr, Nva and ABA with SyrB2. THR, NVA and ABA refer to phosphopantetheinylated Thr, Nva and ABA respectively. NVA is reactive at both the C $^{\gamma}$ and C $^{\delta}$ positions. A truncated form of PPant, upon which the substrate is presented to SyrB2 by SyrB1, is also shown

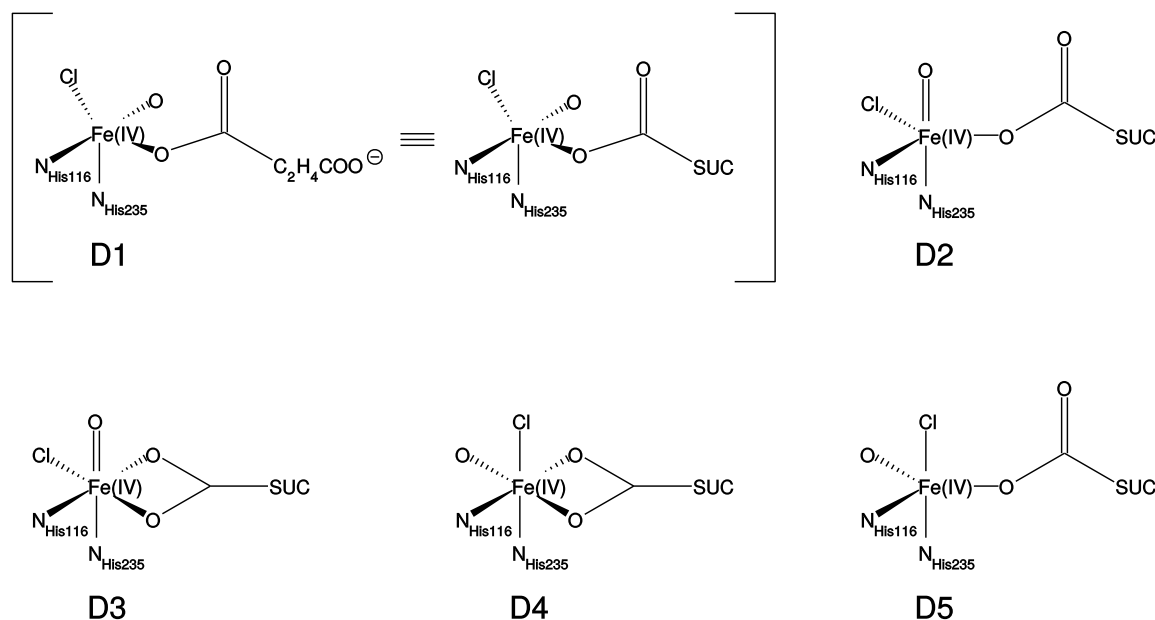


Figure 1.7: Possible isomers of the oxoferryl reactive intermediate

widely [29, 30] believed to be hexacoordinate, with the Fe=O bond *trans* to His235 (species **D3**, Figure 1.7). Later, in explanation of the results of a Mössbauer spectrum that showed the presence of two oxoferryl species [23], an equilibrium between **D3** and **D4** was proposed [31], facilitated by the temporary breaking of an iron-succinate bond. The study calculated this isomerism to have a relatively low barrier. Most recently [26], nuclear resonance vibrational spectroscopy (NRVS) data strongly suggested that the oxoferryl species was in fact pentacoordinate. This was supported by a computational study of both oxygen activation and the following hydrogen abstraction and radical rebound steps, which showed that, depending on the bound substrate, either species **D1** or **D2** was preferred. This study explained the multiple Mössbauer signals of the oxoferryl species through multiple hydrogen-bonding environments.

Whilst many computational studies have attempted to explain SyrB2's absolute selectivity, to date, although many show qualitative preferences for the correct order of selectivity, none reproduce it exactly. This thesis attempts to solve this problem, as well as to resolve the unknown structure of the active-site iron complex, through comparative docking, molecular dynamics, and QM/MM studies of SyrB2 and the three substrates THR, NVA and ABA.

Chapter 2

Theoretical Background

2.1 Overview

One of the central endeavours of computational chemistry is the calculation of the potential energy of a molecule as a function of its nuclear coordinates and electronic state. Once this is possible, other properties, such as equilibrium geometry or energy barriers to a reaction, can be calculated. Whilst the electronic structure of a chemical system is described exactly by the Schrödinger equation (Equations 2.1 and 2.2), for systems containing more than one electron this cannot be solved analytically. As such, computational chemistry techniques are perhaps best divided by the approximations that they make. To so classify the techniques used in this thesis, broadly speaking, *ab initio* techniques approximate the motion of electrons, whilst forcefield methods begin their approximations at the atomic level.

This chapter outlines the techniques that are used over the course of this thesis: density functional theory, coupled-cluster theory and molecular mechanics, as well as their applications in molecular dynamics, docking, hybrid calculations, and optimisation of ground and transition state geometries.

2.2 Quantum Chemical Methods

2.2.1 Overview and the Hartree-Fock Method

Quantum chemical techniques make use of the time-independent Schrödinger equation [32, 33]:

$$\hat{H}\Psi = E\Psi \tag{2.1}$$

Where Ψ is the many-electron wavefunction, E is the energy and \hat{H} is the Hamiltonian operator:

$$\hat{H} = - \underbrace{\sum_i \frac{1}{2} \nabla_i^2 - \sum_k \frac{1}{2m_k} \nabla_k^2}_{\text{Kinetic Terms}} - \underbrace{\sum_i \sum_k \frac{Z_k}{r_{ik}} + \sum_{i < j} \frac{1}{r_{ij}} + \sum_{k < l} \frac{Z_k Z_l}{r_{kl}}}_{\text{Electrostatic Terms}} \quad (2.2)$$

Here i and j refer to electrons, k and l to nuclei, m to mass, Z to nuclear charge, r_{ab} to the distance between a and b , and ∇^2 to the Laplacian operator $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$. In order to calculate the electronic energy, however, an eigenfunction for Equation 2.2 is required, corresponding to a wavefunction describing the electronic motion. Normally nuclear motion is so slow by comparison to electronic motion that neglecting it entirely is not unreasonable (this is known as the Born-Oppenheimer approximation.)

The family of quantum-mechanical techniques is diverse. One of the simplest is the Hartree-Fock (HF) method. In the HF model, the molecular orbitals ϕ are constructed from basis sets of atomic orbitals χ (Equation 2.3). These basis sets can have several forms, varying depending on the application and the software package. One such choice is the Slater-type orbital (STO). These have the form $Nr^{n-1}e^{-\zeta r}$ (where N is a normalising constant, n the principle quantum number, r the distance from the nucleus, and ζ the effective nuclear charge). Whilst these give an excellent representation of the radial part of atomic orbitals, their use comes with a high computational cost. As a result, the most commonly used basis sets for molecular systems are combinations of Gaussian functions, which have the form $e^{-\alpha r^2}$ (where α is a constant). Whilst Gaussian functions do not, individually, resemble atomic orbitals, a combination of several can be used to approximate an orbital. Although reaching a given level of accuracy requires a larger number of Gaussian functions than STOs, in practice the former option normally proves less computationally expensive. Each atomic orbital ϕ_a is then constructed as a linear combination of the M atomic orbitals χ , whose contributions to that orbital are weighted by the orbital coefficients c_{ai} .

$$\phi_a = \sum_{i=1}^M c_{ai} \chi_i \quad (2.3)$$

In the HF method the orbital coefficients c_{ai} of the atomic orbitals χ_i are initially unknown, but are approximated with an iterative procedure. Firstly, an initial “guess” wavefunction is constructed. Due to the variational principle, which states that any trial wavefunction must have an energy equal to or greater than the ground state, the ground state can then be approached by optimising the orbital coefficients to minimise the energy. The energy of each electron is approximated using the one-electron Fock operator (shown in Equation 2.4).

$$\hat{F}(\phi_k(x_1)) = -\frac{1}{2}\nabla^2(x_1) - \sum_k^{\text{nuclei}} \frac{Z_k}{r_{1k}} + \sum_{j=1}^N \left(\hat{J}_j(\phi_k(x_1)) - \hat{K}_j(\phi_k(x_1)) \right) \quad (2.4)$$

Here N refers to the total number of electrons and x_i denotes both the spin and spatial coordinates of the electron i . The Coulombic and exchange interactions between electrons x_1 and x_2 are calculated with the Coulomb operator \hat{J} (Equation 2.5) and the exchange operator \hat{K} (Equation 2.6), respectively.

$$\hat{J}_j(\phi_k(x_1)) = \left(\frac{\int \phi_j^*(x_2) \phi_j(x_2) dx_2}{r_{12}} \right) \phi_k(x_1) \quad (2.5)$$

$$\hat{K}_j(\phi_k(x_1)) = \left(\frac{\int \phi_j^*(x_2) \phi_k(x_2) dx_2}{r_{12}} \right) \phi_j(x_1) \quad (2.6)$$

Once the orbital coefficients are optimised, and consequently the orbitals' energies minimised, the many-electron wavefunction Ψ can be constructed and the energy calculated. The many-electron wavefunction Ψ is approximated as a single Slater determinant of the molecular orbitals.

$$\Psi(x_1, x_2, \dots, x_n) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(x_1) & \phi_2(x_1) & \dots & \phi_n(x_1) \\ \phi_1(x_2) & \phi_2(x_2) & \dots & \phi_n(x_2) \\ \dots & \dots & \dots & \dots \\ \phi_1(x_n) & \phi_2(x_n) & \dots & \phi_n(x_n) \end{vmatrix} \quad (2.7)$$

This yields a wavefunction that is antisymmetric with respect to the exchange of two electrons, a requirement of the Pauli exclusion principle.

$$\Psi(x_1, x_2, \dots, x_n) = -\Psi(x_2, x_1, \dots, x_n) \quad (2.8)$$

The major drawback of the HF method is that it assumes that the motion of the electrons is not correlated, which in real systems is often very far from the truth. Electron correlation can be divided into dynamic and static correlation. Dynamic correlation refers to electrons altering their regular motion to avoid configurations with short interelectronic distances. Static correlation refers to the motion of electrons in systems with a low-lying non-equivalent resonance form, in which one or more electron pairs has been uncoupled. In such systems, a single Slater determinant is not a good reference from which to commence with orbital optimisations, as this leads to a false ground state that is a mixture of the two resonance structures [34]. Such systems are better described by “multi-reference” techniques that use more than one Slater determinant to describe the wavefunction, but these are often highly computationally demanding. Hartree-Fock theory, which ignores both types of correlation, is as a result often unacceptably inaccurate, and sees little

application except as a component of other techniques such as coupled-cluster theory or density functional theory.

2.2.2 Coupled Cluster Theory

Coupled cluster theory builds on the Hartree-Fock wavefunction by calculating and including a correction to it to account for dynamic correlation. This correction considers electronic excitation to a specified degree—typically all single excitations and all pairs of excitations, but much higher orders are often used—by means of a Taylor expansion (Equation 2.9).

$$\Psi_{\text{CC}} = e^{\hat{T}} \Psi_{\text{HF}} = (1 + \hat{T} + \frac{\hat{T}^2}{2!} + \frac{\hat{T}^3}{3!} + \dots) \Psi_{\text{HF}} \quad (2.9)$$

Here the cluster operator, \hat{T} , is the sum of the excitation operators to a given level.

$$\hat{T} = (\hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots + \hat{T}_{N_{\text{elec}}}) \quad (2.10)$$

It is these excitation operators that describe all individual excitations at a given level. For example, the contribution of all possible single excitations are described by the \hat{T}_1 operator, with the contribution of each individual excitation weighted by its t amplitude.

$$\hat{T}_1 \Psi_{\text{HF}} = \sum_i^{\text{occ}} \sum_a^{\text{vir}} t_i^a \phi_i^a \quad (2.11)$$

Here ϕ_i^a is the excitation of an electron from occupied orbital ϕ_i to virtual orbital ϕ_a , and t_i^a the amplitude of this excitation. Substituting Equation 2.10 into Equation 2.9 gives a more complete form of the coupled cluster wavefunction (Equation 2.12)

$$\Psi_{\text{CC}} = e^{\hat{T}} \Psi_{\text{HF}} = (1 + \hat{T}_1 + (\hat{T}_2 + \frac{\hat{T}_1^2}{2}) + (\hat{T}_3 + \hat{T}_2 \hat{T}_1 + \frac{\hat{T}_1^3}{6}) + \dots) \Psi_{\text{HF}} \quad (2.12)$$

Coupled cluster theory is generally referred to by the number of terms calculated as in Equation 2.10. For example, CCSD refers to Coupled Cluster, with Single and Double excitations considered. In addition to the terms calculated explicitly, it is possible to include perturbative corrections to the wavefunction to account for higher orders of excitation. These are denoted by parentheses. For example, CCSD(T) refers to Coupled Cluster, with Single and Double and perturbatively approximated Triple excitations.

Due to the great computational expense and slow basis-set convergence of coupled cluster theory, several methods exist for supplementing the basis level. This thesis makes use of the F12 approximation [35], which, by calculating two-electron functions explicitly, is able to accelerate basis-set convergence.

CCSD(T) is highly accurate for most systems, and is often [36, 37] used as a benchmark for energetic calculations. It is not, however, without its limitations. As it relies on a wavefunction generated by Hartree-Fock theory, it is ultimately a single-reference method, and as such suffers problems with systems with high levels of static correlation. Whilst higher orders of coupled cluster theory would ultimately be able solve the problem through excitation corrections, these are often prohibitively computationally expensive. A number of diagnostics exist to determine the degree to which a system suffers from static correlation.

One means of assessing multireference character is to consider the singles amplitudes. The T_1 diagnostic [38] (Equation 2.13) is a measure of the singles amplitudes scaled across the entire molecule. Whilst useful, it can be misleading if static correlation is localised to a small region of the system [34]. It is calculated as the norm of the \mathbf{t}_1 vector (the vector of the t_i^a amplitudes), weighted by the number of electrons.

$$T_1 = \frac{|\mathbf{t}_1|}{\sqrt{N}} \quad (2.13)$$

The D_1 diagnostic (Equation 2.14) is a measure of the highest singles amplitude. It is the matrix 2-norm of the singles amplitudes matrix $[t_i^a]$ (the largest Euclidean norm of the vectors formed by multiplication of $[t_i^a]$ by a unit vector). This can be useful for systems that are more heterogeneous in their multireference character.

$$D_1 = ||[t_i^a]||_2 \quad (2.14)$$

In addition to these two diagnostics, others are used. A more direct test for multireference character is the value of the highest doubles amplitude t_{ij}^{ab} [39]. As a resonance form with an uncoupled electron pair *is* a double excitation, this is, by definition, always large in a system with a high level of static correlation. It is, however, a somewhat insensitive diagnostic in borderline cases [34]. The T_1/D_1 ratio gives an idea of the homogeneity of the multireference character of the system [40]; a system of perfectly homogeneous multireference character will have a T_1/D_1 ratio of $\frac{1}{\sqrt{2}}$. Recently, a new diagnostic, A_λ , based purely on DFT calculations, has been proposed [41]. This is defined in equation 2.15. It utilises the ratio of total energy of atomisation calculated by a pure functional ($\Delta_{at}E(0)$) and by a functional with $\lambda\%$ Hartree-Fock exchange ($\Delta_{at}E(\lambda)$) (see section 2.2.3).

$$A_\lambda = \frac{(1 - \Delta_{at}E(\lambda))/\Delta_{at}E(0)}{\lambda} \quad (2.15)$$

2.2.3 Density Functional Theory

Overview

Density functional theory [42, 43], for whose development Walter Kohn shared the 1998 Nobel prize, ignores the many-electron wavefunction entirely, employing instead the electron density. The density has numerous advantages over the wavefunction, not least the fact that it is experimentally measurable. As with the Hartree-Fock method, molecular orbitals are constructed with LCAO-MO theory from a basis set of atomic orbitals. The electron density is then constructed from these molecular orbitals.

$$\rho = \sum_i^N |\phi_i^{KS}|^2 \quad (2.16)$$

The energy is minimised with respect to the density. It is at a minimum when the Kohn-Sham equations are fulfilled:

$$\hat{h}^{KS} \phi_i^{KS}(x) = \epsilon_i^{KS} \phi_i^{KS}(x) \quad (2.17)$$

Here \hat{h}^{KS} refers to the Kohn-Sham single-electron Hamiltonian

$$\hat{h}^{KS} = -\frac{1}{2}\nabla^2 + \hat{v}_{ext}(x_1) + \int dx_2 \frac{\rho(x_2)}{r_{12}} + \hat{v}_{xc}(x_1) \quad (2.18)$$

Where

$$\hat{v}_{ext}(x_1) = \sum_k^{nuclei} \frac{Z_k}{r_{1k}} \quad (2.19)$$

and

$$\hat{v}_{xc}(x_1) = \frac{\partial E_{xc}(\rho)}{\partial \rho} \quad (2.20)$$

From the density, the energy can be calculated.

$$E_{KS}[\rho] = T_S[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \quad (2.21)$$

Here $T_S[\rho]$ is the kinetic energy, $E_{ne}[\rho]$ the electrostatic attraction between the nuclei and the electrons, $J[\rho]$ the electron-electron repulsion (also on the assumption of non-interacting electrons), and $E_{xc}[\rho]$ the exchange-correlation energy. $E_{ne}[\rho]$ is calculated under the Born-Oppenheimer approximation. $T_S[\rho]$ and $J[\rho]$ are calculated by assuming the absence of correlation in the electron's movements. The kinetic and Coulombic energy due to correlated electronic motion is approximated in the $E_{xc}[\rho]$ term, which also includes spin interactions.

$$E_{xc}[\rho] = \underbrace{T[\rho] - T_S[\rho]}_{\text{Kinetic Contribution}} + \underbrace{E_{ee}[\rho] - J[\rho]}_{\text{Exchange and Correlation Contribution}} \quad (2.22)$$

Here the kinetic correlation energy is expressed as the difference between the true kinetic energy $T[\rho]$ and the kinetic energy calculated on the assumption of non-interacting electrons $T_S[\rho]$. This is generally small. The exchange energy, $E_{ee}[\rho]$, refers to the energy due to electron-electron interactions, from both spin and correlation contributions.

Exchange-Correlation Approximations

Although it has been shown by the Hohenberg-Kohn theorem that the energy can be calculated exactly from the density, the form of the functional that does so is not known, and currently exchange and correlation can only be approximated. A number of approximate functionals have been developed, which tend to vary in reliability depending on the system in question.

- The Local Density Approximation (LDA) is the approximation that the electron density around each point is uniform, making $E_{xc}[\rho]$ a function, rather than a functional, of the density at that point. The form of $E_{xc}[\rho]$ for a uniform electron density is calculable using quantum Monte Carlo calculations, making the LDA highly accurate within the bounds of this assumption. The assumption is rather an audacious one, however, due to the significant heterogeneity of the electron density of real chemical systems. For this reason, the LDA sees most of its practical use as a component of more sophisticated functionals.
- The Generalised Gradient Approximation (GGA) is that $E_{xc}[\rho]$ at a given point is a functional of the density and its gradient at a given point. The family of GGA functionals is quite diverse, with a large number of methods available for the incorporation of the gradient. B88 [44], for example, adds a gradient contribution to the LDA energy, whilst PW86 [45] includes it as a scaling factor.
- Hybrid functionals use contributions from multiple components to construct the exchange energy. Since electron exchange can be calculated exactly in Hartree-Fock theory (Equation 2.4), doing so and adding this as a contribution to the DFT energy might intuitively appear to be the ideal method for calculating the exact exchange in DFT. Unfortunately, mixing absolute values of contributions to the total energy calculated with these two techniques tends to lead to problems, meaning it is normally preferable to include only a fraction of the Hartree-Fock exchange, with the rest approximated with a density functional approach. As Hartree-Fock exchange

stabilises higher multiplicities, a hybrid functional will favour higher spin-states in proportion to the amount of exact exchange it employs [46]. Many hybrid functionals have been proposed, with B3LYP (Equation 2.23) being probably the most widely known and used:

$$E_{xc}^{B3LYP} = (1 - a)E_x^{LDA} + aE_x^{HF} + b\Delta E_x^{B88} + (1 - c)E_c^{LDA} + cE_c^{LYP} \quad (2.23)$$

Where E^{B88} , E^{LYP} , E^{HF} and E^{LDA} are the energies calculated using B88, LYP [47], Hartree-Fock theory and the local density approximation respectively, and the subscripts x and c referring respectively to the exchange and correlation components.

Empirical Dispersion Corrections

DFT suffers problems with the treatment of London dispersions [48]. Numerous schemes have arisen to compensate for this, including empirical corrections, correction potentials, and parameterised functionals [49]. This thesis makes use of the DFT-D3 correction of Grimme *et al.* [50]. This is an empirical correction to the DFT energy

$$E_{DFT-D3} = E_{DFT} - E_{D3} \quad (2.24)$$

Where E_{D3} is the D3 dispersion correction. This is the sum of two- and three- body terms.

$$E_{D3} = E^{(2)} - E^{(3)} \quad (2.25)$$

E_{D3} is dominated by the two body term

$$E^{(2)} = \sum_{AB} \sum_{n=6,8,10} s_n \frac{C_n^{AB}}{r_{AB}^n} f_{d,n}(r_{AB}) \quad (2.26)$$

Where C_n^{AB} are the n^{th} -order dispersion coefficients, s_n is a scaling factor, r_{AB} is the distance between the atoms A and B , and $f_{d,n}(r_{AB})$ is a damping function, which prevents problems with very small internuclear distances.

2.3 Molecular Mechanics

Molecular mechanics (MM) refers to calculations that use atomic-level parameters, which allows the use of classical physics. Such parameters include atomic point charges to describe electrostatic interactions, or force constants to describe covalent bond stretches using Hooke's law. Whilst this means that these parameters must be derived by other

means and remain fixed, MM calculations are faster than QM calculations by several orders of magnitude, and can be run on much larger chemical systems. As a result, MM calculations see wide use in the study of intermolecular interactions of large systems such as proteins, which are dominated by terms such as dipole-dipole interactions and London dispersions. Generally speaking, the MM energy is the sum of the energies from bond stretches, angle and torsional distortion, electrostatic interactions and London dispersions (equation 2.27).

$$E_{\text{MM}} = E_{\text{Bond}} + E_{\text{Angle}} + E_{\text{Torsion}} + E_{\text{Elec}} + E_{\text{London}} \quad (2.27)$$

More specifically, bonds and angles are often represented using harmonic potentials, electrostatics using Coulomb’s law, and London dispersions using a Lennard-Jones potential. A simple forcefield is shown in Equation 2.28.

$$E_{\text{MM}} = \sum_{\text{Bonds}} k_b(d - d_{eq})^2 + \sum_{\text{Angles}} k_\theta(\theta - \theta_{eq})^2 + \sum_{\text{Torsions}} k_\phi[1 + \cos(n\phi + \delta)] \\ + \sum_{AB} \left\{ \underbrace{\frac{q_A q_B}{4\pi\epsilon_0 r_{AB}}}_{\text{Electrostatics}} + \underbrace{4\epsilon_{AB} \left[\left(\frac{\sigma_{AB}}{r_{AB}} \right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}} \right)^6 \right]}_{\text{London Dispersions}} \right\} \quad (2.28)$$

Here d , θ and ϕ refer to bonds, angles and dihedrals and A and B to nonbonded atoms, respectively. This requires a set of values for the parameters: equilibrium bond lengths and angles (d_{eq} and θ_{eq}), the constants k_b , k_θ and k_ϕ , the dihedral phase angle δ , the Lennard-Jones well depths and point of zero potential energy (ϵ and σ respectively), and the charges (q).

Whilst the results of a molecular mechanics calculation are only accurate for the electronic state set in the parameters, many properties of larger systems (such as proteins) are dominated by these interactions. This makes molecular mechanics particularly useful in such applications.

Some minor modifications to a standard MM forcefield of the type shown in Equation 2.28 are often used to speed up calculations:

- London dispersion interactions are truncated by applying simple cutoffs to the Lennard-Jones potentials.
- For calculations involving explicit solvent, periodic boundary conditions, or PBC, are often employed. These address the problem that would arise at the edge of the solvent shell, by treating the system as being infinitely repeating in each direction.

- Where PBC are employed, Coulombic interactions are usually calculated using the particle-mesh Ewald method. This applies a cut-off to the electrostatic interactions experienced by each particle, beyond which they are treated as a sum in Fourier space.

2.4 Hybrid QM/MM calculations

Hybrid quantum mechanical/molecular mechanical (QM/MM) methods couple a QM calculation run on one region of a chemical system to an MM calculation run on the rest of it. They are particularly applicable to the modelling of reactions in systems such as enzymes, in which only a small number of atoms is significant electronically but a much larger number is sterically important. Since their inception in the 1970s [51, 52], QM/MM calculations have taken such a prominent place in the computational chemist’s toolkit that the 2013 Nobel prize was awarded to Martin Karplus, Michael Levitt and Arieh Warshel for their development.

In a QM/MM calculation, all atoms of the system \mathbf{S} are assigned uniquely to either the QM region \mathbf{Q} or the MM region \mathbf{M} [53, 54]. In most QM/MM schemes, “link” atoms, typically hydrogen, are added to \mathbf{Q} to atoms that have a bond to an atom in \mathbf{M} . One can define the QM/MM energy as the sum of the QM and MM energies, as well as a term E_{Int} , which covers interactions between the two regions (Equation 2.29).

$$E_{QM/MM}(\mathbf{S}) = E_{QM}(\mathbf{Q}) + E_{MM}(\mathbf{M}) + E_{Int}(\mathbf{Q}, \mathbf{M}) \quad (2.29)$$

E_{Int} contains contributions from electrostatic interactions, London dispersions and bonded interactions (Equation 2.30)

$$E_{Int}(\mathbf{Q}, \mathbf{M}) = E_{elec}(\mathbf{Q}, \mathbf{M}) + E_{London}(\mathbf{Q}, \mathbf{M}) + E_{bond}(\mathbf{Q}, \mathbf{M}) \quad (2.30)$$

$E_{London}(\mathbf{Q}, \mathbf{M})$ is normally calculated using a simple Lennard-Jones potential. This requires parameters for the QM atoms, which might appear challenging as it is not unusual to include more “exotic” atoms with unusual properties within \mathbf{Q} (for example transition metal centres of metalloenzymes). In practice it is rarely a problem, however, as Lennard-Jones potentials are significant over such a short distance that they normally only affect atoms at the boundary, and it is generally advisable (in order to avoid misrepresenting the electronic structure), to place the boundary away from atoms suspected of having unusual properties. $E_{bond}(\mathbf{Q}, \mathbf{M})$ is normally calculated at the MM level.

There are several methods for calculating $E_{elec}(\mathbf{Q}, \mathbf{M})$. One can, for example, calculate $E_{MM}(\mathbf{M})$ in the presence of point charges representing \mathbf{Q} , a scheme known as mechanical

embedding. This has some significant limitations. Firstly, it means that \mathbf{M} is entirely absent during the calculation of the $E_{QM/MM}(\mathbf{Q})$, requiring the daring assumption that the electronic structure of \mathbf{Q} is not polarised by the environment. Secondly, it requires reliable point charges for \mathbf{Q} to be calculated each time the electronic structure changes, which often proves challenging.

For these reasons, the use of electrostatic embedding is far more common. In this scheme, $E_{elec}(\mathbf{Q}, \mathbf{M})$ is calculated at the QM level, by calculating $E_{QM}(\mathbf{Q})$ in the presence of point charges representing \mathbf{M} . Whilst this scheme is generally preferable to mechanical embedding, it is not without its limitations. Whilst the charge density of \mathbf{Q} can be polarised by \mathbf{M} , the reverse is not true. Consequently, \mathbf{Q} may become artificially polarised to compensate, particularly at the boundary. Whilst the error this causes is usually acceptably low, the problem becomes far worse at cross-boundary bonds, which, in a link atom scheme, will have a link atom of \mathbf{Q} separated from a point charge of \mathbf{M} by a distance of less than a bond length. There are several solutions to this problem. One can simply delete the partial charges of atoms in \mathbf{M} that have a cross-boundary bond during QM calculations. This, however, creates an artificial polarisation of the opposite charge. It is usually preferable to shift the point charges of such atoms to adjacent atoms in \mathbf{M} , a scheme known as charge shifting. This mitigates the over-polarisation within \mathbf{Q} , whilst preserving the net charge, and even some of the boundary polarisation, of \mathbf{M} .

2.5 Geometry Optimisation

With the ability to calculate the potential energy as a function of a given set of nuclear coordinates comes access to a potential energy surface relating energy to geometry. Perhaps the most useful application of this is the determination of the most stable molecular structure, which corresponds to a minimum in the potential energy surface. Whilst finding the global minimum is mathematically incredibly difficult due to the vast numbers of local minima on the potential energy surface, with human oversight this is often not a problem. It is, in fact, often more useful to calculate the minimum energy of a given conformer or isomer, rather than the minimum energy structure with a given empirical formula.

Finding a minimum on a potential energy surface can be done iteratively using the Newton-Raphson method (Equation 2.31), which uses the first and second derivatives of the energy with respect to the coordinates at each point n to scale the value of the next trial point $n + 1$.

$$\mathbf{q}_{n+1} = \mathbf{q}_n - (\mathbf{H}_n)^{-1} \mathbf{g}_n \quad (2.31)$$

Here \mathbf{q} is a coordinate vector, \mathbf{g} a gradient vector (a vector of the derivatives of the energy with respect to the coordinates of each atom) and \mathbf{H} a Hessian matrix of the second derivatives of the energy with respect to coordinates.

Construction of the inverse Hessian matrix \mathbf{H}^{-1} can be extremely computationally costly, particularly for larger systems, as the Hessian has $3N \times 3N$ elements. For this reason, most modern optimisation algorithms use a variant of the Newton-Raphson scheme that allows \mathbf{H} to be updated approximately rather than constructed *de novo* at each step. The initial elements of \mathbf{H} can be calculated prior to the first step, or can simply be a unit matrix. Examples of approximate update schemes include Broyden-Fletcher-Goldfarb-Shanno (BFGS) [55, 56, 57, 58] and GEDIIS [59]. For large molecular systems, the full \mathbf{H} can become so large that storing it in memory becomes infeasible. The limited-memory variant of the BFGS algorithm (L-BFGS) was designed for such systems [60, 61]. This updates the BFGS scheme at each step for a given number of previous steps, rather than for the entire optimisation, which can greatly reduce computational expense.

In addition to the calculation of equilibrium geometries, it is often desirable to calculate transition state geometries and energies. These correspond to saddle points in the potential energy surface. In this thesis, these were calculated with the partitioned rational function optimisation (P-RFO) method [62]. Firstly, an initial Hessian matrix of displacements from the starting geometry is calculated. At this point, if the structure is near a transition state, one direction will lead to a lower energy in all but one dimension. This direction will lead to the saddle point.

During QM/MM calculations, the majority of the computational expense is consumed by the QM calculations. Consequently, it is preferable to reduce the number of QM calculations, even at the cost of an increased number of MM calculations. This is accomplished by microiterative optimisation schemes, which optimise an inner region as described above, but, following each step, freeze the coordinates of this inner region and fully optimise the rest of the molecule around it.

Another scheme that can reduce computational expense is the use of Hybrid Delocalised Internal Coordinates (HDLCs) [63]. This scheme divides the protein into residues, each containing a set of atoms in primitive internal coordinates. This allows Cartesian steps, which optimise the positions of the residues relative to one another, to be interspersed with HDLC steps, which optimise the relative positions of the atoms.

2.6 Molecular Dynamics

Since it is possible to calculate the forces acting on each atom of a molecular system, it is also possible, through Newton’s laws of motion, to simulate thermal motion. This

can illuminate interesting features of chemical systems, and today molecular dynamics (MD) simulations see wide use in calculations of the properties of biomolecules, for which dynamic properties often contribute to functionality [64]. Whilst ab-initio molecular dynamics simulations are seeing increasing use, this thesis deals exclusively with MD using molecular mechanics for energy calculations. Again, these typically use periodic boundary conditions to avoid problems resulting from interactions with an artificial surface, and Ewald summation to account for Coulombic interactions in the periodic system.

The numerical integration of Newton’s equations of motion, to determine position as a function of time, can use one of several integrators. This thesis makes use of the leapfrog algorithm, which couples the propagation of the position x (Equation 2.32) and velocity v (Equation 2.33. Here a refers to acceleration).

$$x(t + \Delta t) = x(t) + v(t + \frac{\Delta t}{2})\Delta t \quad (2.32)$$

$$v(t + \frac{\Delta t}{2}) = v(t - \frac{\Delta t}{2}) + a(t)\Delta t \quad (2.33)$$

MD simulations are classified according to which system variables are conserved. Perhaps the simplest arrangement is the NVE, or microcanonical, ensemble, which conserves number of particles, volume and energy. In this ensemble, the potential energy and kinetic energy can be interchanged, but the overall energy must be maintained. In simulations of chemical systems, however, the total energy is often a far less interesting system variable than the pressure and temperature, so MD simulations are often run under other ensembles.

One can keep a closed system under constant pressure by controlling the temperature or vice versa, leading to several alternative classes of statistical mechanical ensembles. This thesis uses two types: NVT, or canonical, ensembles, which have constant volume and temperature but allow pressure to fluctuate, and NPT, or isothermal-isobaric, ensembles, which have constant pressure and temperature but allow volume to fluctuate.

In the case of a canonical ensemble, one requires a molecular “thermostat” to maintain the temperature. In this thesis, this is done by means of the v -rescale thermostat [65], which re-scales the velocity of each atom by a factor α derived from the ratio of the target kinetic energy K_t (which is derived stochastically from the target temperature) and the current kinetic energy K (Equation 2.34).

$$\alpha = \sqrt{\frac{K_t}{K}} \quad (2.34)$$

Similarly, for an isothermal-isobaric ensemble one requires a molecular “barostat”—this thesis makes use of the procedure of Berendsen [66], in which cell parameters are scaled

by a factor μ to expand or compress the system to a desired pressure.

$$\mu = \left[1 - \frac{\Delta t}{\tau_p}(P - P_t) \right]^{1/3} \quad (2.35)$$

Where P_t is the target pressure, P is the current pressure, Δt is the timestep, and τ_p is a coupling constant.

2.7 Docking

Docking algorithms attempt to ascertain the affinity and conformation of the binding of a ligand to a receptor. Although this can more generally be applied to protein-protein interactions, the scope of docking within this thesis is limited to the placement of substrate within SyrB2’s active site channel. This task can be broadly divided into two subtasks: calculating the affinity of a given pose of the ligand within the protein receptor, and the generation and optimisation of such poses. There are many, qualitatively different, approaches to both of these problems. To make the latter problem tractable, bonds are typically treated as fixed in length, but rotation is permitted around user-specified bonds.

Due to the size of the substrate-pantetheinyl complex, Autodock VINA was chosen for this study, as it has been shown to be able to reliably handle large numbers of rotatable bonds [67, 68]. In tests on the PDBbind data set [68] of 190 known protein-ligand complexes, VINA showed very high speeds, around 62 times faster than Autodock 4. This did not come at the compromise of accuracy: Autodock VINA was able to acquire an RMSD of under 2Å in around 78% of cases, comparing favourably to Autodock 4 which managed this in only 49%, with particularly striking differences in accuracy for complexes with more rotatable bonds.

Autodock VINA calculates affinity with an empirical scoring function, which is applied to each pair of atoms that can move relative to one another. This is a weighted sum of six terms (shown in Table 2.1), representing hydrophobic interactions, hydrogen bonding, the number of the rotatable bonds between the atoms in question, and three steric terms. As electrostatics are ignored, partial charges are not used. Pose optimisation is carried out by the iterated local search global optimiser, which performs a series of mutations and subsequent local optimisations. For each pose the gradient, as well as the value, of the scoring function is calculated, which speeds up pose optimisation.

Table 2.1: Terms included in the Autodock VINA scoring function. Here d is the inter-atomic separation

Term	Weight	Form
gauss ₁	-0.0356	$e^{-(d/0.5 \text{ \AA})^2}$
gauss ₂	-0.00516	$e^{-((d-3 \text{ \AA})/2 \text{ \AA})^2}$
Repulsion	0.840	d^2 , if $d < 0$ 0, if $d \geq 0$
Hydrophobic	-0.0351	1, if $d < 0.5 \text{ \AA}$ 0, if $d > 1.5 \text{ \AA}$ (with linear interpolation in between)
Hydrogen bonding	0.587	1, if $d < -0.7 \text{ \AA}$ 0, if $d > 0 \text{ \AA}$ (with linear interpolation in between)
N _{rot}	0.0585	

Chapter 3

Electronic Structure of SyrB2's Oxoferryl Intermediate

3.1 Introduction

As discussed in Chapter 1, the Fe(IV)-oxo unit was discovered in proteins as early as the 1980s, and has been studied extensively, both experimentally and computationally, in both proteins [25, 69, 70, 71] and inorganic complexes [36, 72] (for more comprehensive reviews see [73] and [74]). From these studies, it has emerged that the unit's electronic structure is not easily represented computationally. As a result of the near-degenerate d-orbitals of the iron centre, the system has complicated spin-state energetics, a problem not handled well by density functional theory.

In spite of this deficiency of density functional theory, due to the large number of atoms involved in enzyme reactions the computational chemist is often left with no better alternative method. As a result, the limitations of DFT when applied to Fe(IV)-oxo structures have been the subject of intense study and debate. SyrB2's oxoferryl intermediate **D** is known from Mössbauer spectroscopy [27] to have a quintet ground state, but is also believed to have low-lying triplet and septet states. Model **1a** (shown in Figure 3.3), which is a model complex for species **D3**, has been studied at the CASPT2 level by the group of Siegbahn [31]. They compared the spin-state energetics calculated at this level to results using B3LYP and B3LYP* (two variants of the B3LYP functional which use 20% and 15% exact exchange, respectively), and found both functionals to give a reasonably similar representation of the spin-state energetics to the CASPT2 data.

Studies of haem systems, in particular Cpd 1 of CytP450, have addressed the application of DFT to oxoferryl species (for a detailed review see [75]). Several studies of P450cam [69, 70, 71] found B3LYP to reproduce the energy separation of excited states

from multireference techniques reasonably well. However, Cpd 1 is quite different from species **D** of SyrB2 electronically—Cpd 1 is formally a perferryl complex, though it is more accurately described as a ferryl complex with a radical porphyrin ligand [76], whilst species **D** of SyrB2 is a true ferryl complex.

The Solomon group [77] studied the oxygen adduct rather than the oxoferryl species, using NO bound to the complex $\text{Fe}(\text{Me}_3\text{TACN})(\text{NO})(\text{N}_3)_2$ (TACN referring to 1,4,7-triazacyclononane) as a model for the oxygen adduct. The study investigated how various functionals reproduced the geometry from a crystal structure and several parameters from spectra, finding B3LYP to give a poor representation of the spectral data but that BP86 with 10% HF exchange gave a more accurate representation.

There have also been attempts to use CCSD(T) as a benchmark for some model complexes. Chen and co-workers [36] tested 29 different functionals against CCSD(T) for a hydrogen abstraction reaction by several oxoferryl model complexes, finding B3LYP and TPSSH to give particularly good results, both for barriers and for the difference between the quintet and triplet barriers. Geng *et al.* [72] also compared CCSD(T) to B3LYP in a similar hydrogen abstraction reaction with the same Fe(IV) model complex, finding reasonable agreement between the two techniques, but noting that they had not yet reached CCSD(T) basis-set convergence.

In this chapter, the reliability with which various density functionals address the electronic structure of Fe(IV)-oxo model complexes is addressed. Firstly, DFT-optimised structures of several inorganic complexes are compared to crystal structures. Next, the relative energies of the quintet, triplet and septet states of a model complex from DFT are compared to those obtained by the CASPT2 calculations of Borowski and co-workers. CCSD(T) is then investigated as a potential benchmark-level technique. Finally, the relative energies of several geometric and structural isomers of the model complex are compared at the three multiplicities.

3.2 Computational Details

DFT calculations, unless otherwise specified, were run using the TURBOMOLE suite of programs [78, 79, 80] at the def2-TZVP [81, 82] basis level. These were carried out in the gas phase, with a D3 empirical dispersion correction [50], using the multipole accelerated resolution of identity approximation [83], at the respective multiplicity. CCSD(T) single-point calculations were carried out using the MOLPRO suite of programs [84, 85, 86, 87] on a structure optimised at the B3LYP/def2-TZVP level, as described above, at the respective multiplicity. Explicit correlation was implemented through the F12 approximation [35]. CCSD(T) calculations used restricted open-shell Hartree-Fock or density functional

theory for calculation of the reference orbitals, with the coupled-cluster part of the calculation performed with the spin unrestricted. RKS/CCSD(T) calculations used the B3LYP functional for calculation of the reference orbitals.

3.3 Results and Discussion

3.3.1 Calculation of Structural Parameters

A crude but important measure of a technique’s reliability is its ability to reproduce equilibrium geometries accurately. This simple test was applied to a number of density functionals using the CSD [88] crystal structures of two oxoferryl complexes, trans-[Fe(IV)(O)(TMC)(NCCH₃)]²⁺ [89] and [Fe(IV)(O)(TMC-py)]²⁺ [90] (TMC referring to 1,4,8,11-tetramethyl-1,4,8,11-tetraazacyclotetradecane, see Figure 3.1). Whilst these models are quite similar to species **D** of SyrB2, it must be remembered that there are significant differences in their ligand environments, which are higher-field than those of species **D**, leading to triplet ground states for both complexes. In addition, the ligands of these complexes are polydentate, which may make them less flexible. It must also be remembered that the crystal structure of a complex will be somewhat distorted from the gas-phase structure, placing an upper limit on the reliability of comparisons such as these. A set of functionals was chosen to cover a range of exact exchange contributions—PBE and BLYP with 0%, B3LYP with 20%, PBE0 with 25% and BHLYP with 50%.

Table 3.1: Structural parameters of [Fe(IV)(O)(TMC)(NCCH₃)]²⁺. N_{eq} refers to any of the four equatorial TMC nitrogen ligands

Structure	Fe=O (Å)	Fe-NCCH ₃	Average Fe-N _{eq} (Å)
Crystal Structure	1.65	2.06	2.09
BHLYP-D	1.72	2.06	2.11
PBE0-D	1.60	2.08	2.10
B3LYP-D	1.62	2.11	2.13
PBE-D	1.64	2.05	2.13
BLYP-D	1.65	2.11	2.18
B3LYP	1.62	2.13	2.15

All functionals gave reasonable geometries for both structures, reproducing the distances to within 0.11 Å or better in each case (see Tables 3.1 and 3.2). It would be dangerous to attempt to compare functionals based on these findings, as all displayed a level of accuracy that approaches the limits of the reliability of this test. However, the

Table 3.2: Structural Parameters of $[\text{Fe(IV)}(\text{O})(\text{TMC-py})]^{2+}$. N_{Py} refers to the nitrogen of the pyridine ligand

Structure	Fe=O (\AA)	Fe- N_{Py} (\AA)
Crystal Structure	1.67	2.12
BHLYP-D	1.73	2.12
PBE0-D	1.61	2.17
B3LYP-D	1.62	2.20
PBE-D	1.65	2.16
BLYP-D	1.65	2.23
B3LYP	1.67	2.12

ability to reproduce equilibrium geometry, as important as it is, does not imply the ability to replicate spin-state energetics.

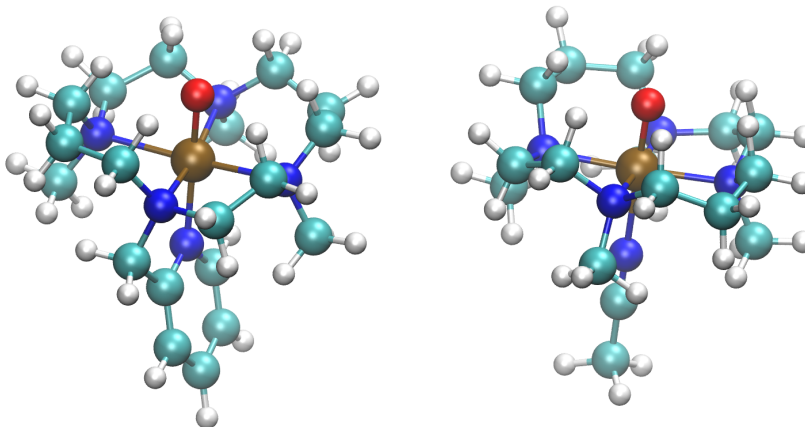


Figure 3.1: $[\text{Fe(IV)}(\text{O})(\text{TMC-py})]^{2+}$ (left) and $[\text{Fe(IV)}(\text{O})(\text{TMC})(\text{NCCH}_3)]^{2+}$ (right)

3.3.2 Electronic Structure of Oxoferryl Complexes

To address the problem of spin-state energetics, a hexacoordinate model complex of SyrB2’s oxoferryl active-site intermediate was considered, with the succinate truncated to acetate and the histidines to imidazoles (Model **1a**, Figure 3.3). This model is the same as that used by Borowski *et al.* [31], allowing spin-state separations to be compared directly to their CASPT2 results. Results are shown in Figure 3.2 and Table 3.3.

A functional will favour a higher multiplicity in proportion to the percentage of exact

exchange it includes (see Section 2.2.3). This is evident in these results, with the triplet rising in energy and the septet falling in energy relative to the quintet ground state almost linearly with percentage exact exchange. The trend is not perfect, however, as the functionals have differences from one another other than exact exchange contribution. All functionals correctly place the quintet as the ground state (in agreement with Mössbauer spectroscopy [27]), although the relative ordering of the excited states is incorrect (by comparison to CASPT2) in PBE0 and B3LYP. The functional that best replicates the CASPT2 orderings is B3LYP, although, as noted by Borowski *et al.*, the septet excited state is placed rather too low in energy by this functional.

Spin-State Ordering by Density Functional

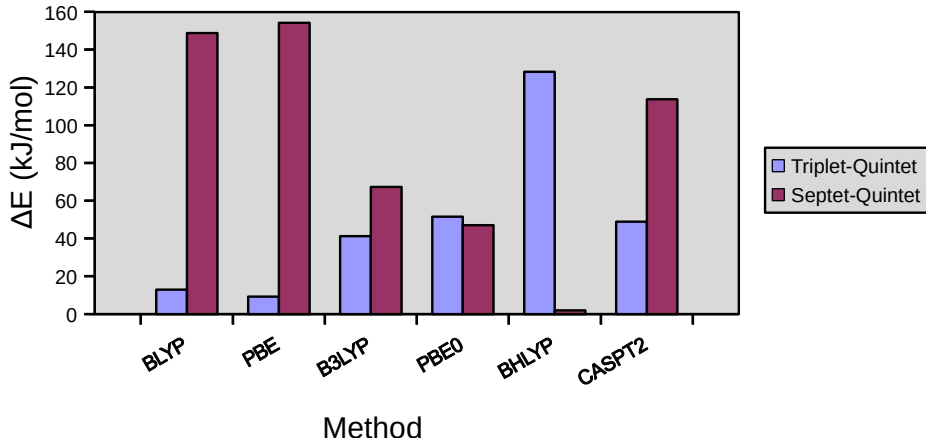


Figure 3.2: Spin-state ordering (relative to quintet state) of Model **1a** by method, in ascending order of exact-exchange contribution.

Table 3.3: Spin-state ordering of Model **1a** (relative to quintet) by density functional

Functional	Quintet-Triplet Separation (kJ/mol)	Quintet-Septet Separation (kJ/mol)
BLYP-D	13	149
PBE-D	9	154
B3LYP-D	41	67
PBE0-D	52	47
BHLYP-D	128	2
CASPT2 [31]	49	114

The valence orbitals of this complex, shown in Figure 3.4, are typical of complexes of this class [72, 91]. In the quintet ground state, the HOMO is a $d_{x^2-y^2}$ orbital of

antibonding character with respect to all xy- plane bonds. Upon moving from the quintet state to the triplet, this orbital is emptied, with a corresponding decrease in all xy-plane Fe–L bond lengths (Figure 3.3). The LUMO is a d_{z^2} orbital of antibonding character with respect to the Fe=O and Fe–N_T bonds, which is filled upon excitation to the septet, resulting in an increase in these bond lengths. Below these sit d_{xz} and d_{yz} type orbitals that are half filled in all spin states and are slightly antibonding in character with respect to Fe=O. The contribution from the ferryl oxygen to these orbitals is evidenced by its spin-populations, which for the triplet and quintet are 0.8 and 0.6 respectively, despite the fact that formally the oxygen bears no unpaired electrons in these spin states. Below these sit the d_{xy} orbitals, which are essentially non-bonding, and below these are the oxygen p_x and p_y orbitals. Upon formation of the septet an electron is promoted from one of these oxygen p orbitals, leaving an unpaired electron behind, leading to the oxygen’s considerably higher spin population in the septet state.

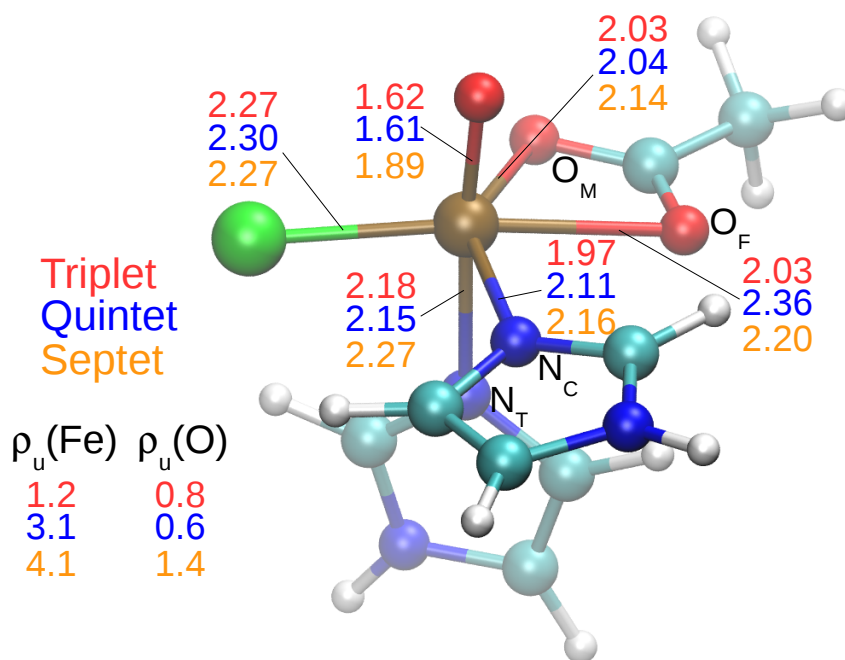


Figure 3.3: Bond lengths and spin densities for Model **1a** in the quintet, triplet and septet spin states calculated at the B3LYP-D/def2-TZVP level. All lengths are given in Ångströms. ρ_u refers to the spin density $\rho_\alpha - \rho_\beta$ from the Mulliken spin population of the atom in question.

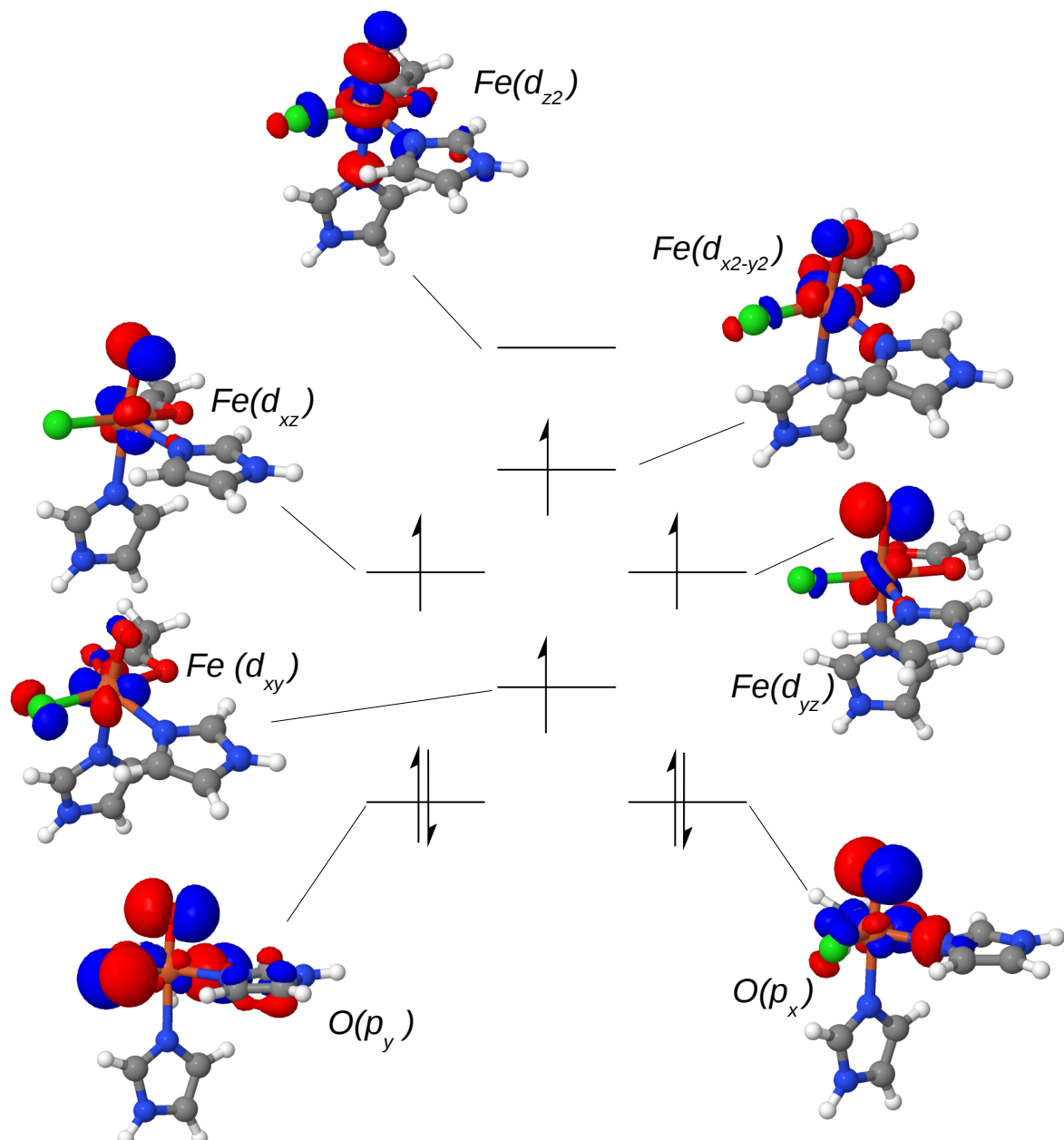


Figure 3.4: Valence orbitals of Model **1a**, shown with quintet multiplicity.

3.3.3 Comparison of DFT to CCSD(T)

Due to the functional dependence of the spin energetics, the ability to run benchmark-level calculations for a given structure would be extremely useful. Whilst CASPT2 calculations such as those of Borowski *et al.* can have a high level of accuracy, this level is entirely dependent on the user's choice of the active space of orbitals. In the hope of obtaining benchmark-level calculations that could be used to validate DFT and compared to CASPT2, CCSD(T) was investigated, as the accuracy of this method is not dependent on user-selected active space and is partly measurable by several diagnostics. Due to the high

computational cost of CCSD(T), a smaller model was used (Model **1b**, see Figure 3.7), in which histidine was represented as ammonia and succinate as formate. This truncation appears not to overly affect electronic structure (see Table 3.4 for descriptors).

Table 3.4: Comparison of models **1a** and **1b** at B3LYP/def2-TZVP level. Bond lengths given at quintet multiplicity.

Parameter	Model 1a	Model 1b
Fe=O (Å)	1.61	1.61
Fe-Cl (Å)	2.30	2.26
Fe-N _T (Å)	2.15	2.16
Fe-N _C (Å)	2.11	2.16
Fe-O _F (Å)	2.36	2.40
Fe-O _M (Å)	2.04	2.05
Quintet-Triplet (kJ/mol)	41	48
Quintet-Septet (kJ/mol)	67	71

Figure 3.5 shows basis set convergence of CCSD(T) calculations using a restricted open-shell Hartree-Fock (RHF) reference wavefunction, both with and without F12 explicit correlation. It would appear that whilst the basis set level is approaching convergence, it has not yet reached it. Figure 3.6 and Table 3.5 show spin-state splittings for these basis set levels. Again, all calculations correctly placed the quintet as the ground state. There is a trend towards an increased basis level giving a lower value of the triplet and a higher value of the septet. However, it is only the very highest basis level (cc-pVTZ with explicit correlation) that reproduces the relative placement of the triplet and the septet of CASPT2, again suggesting that basis-set convergence has not yet been reached.

Table 3.5: Relative energies of spin-states of Model **1b** from CCSD(T). CASPT2 data from literature, calculated on Model **1a** [31], are also shown

Method	Quintet-Triplet (kJ/mol)	Quintet-Septet(kJ/mol)
CASPT2	49	114
ROHF/CCSD(T)/cc-pVDZ	82	45
ROHF/CCSD(T)/cc-pVTZ	69	62
ROHF/CCSD(T)/cc-pVDZ/F12	61	52
ROHF/CCSD(T)/cc-pVTZ/F12	64	66
RB3LYP/CCSD(T)/cc-pVDZ	73	62
RB3LYP/CCSD(T)/cc-pVTZ	61	80

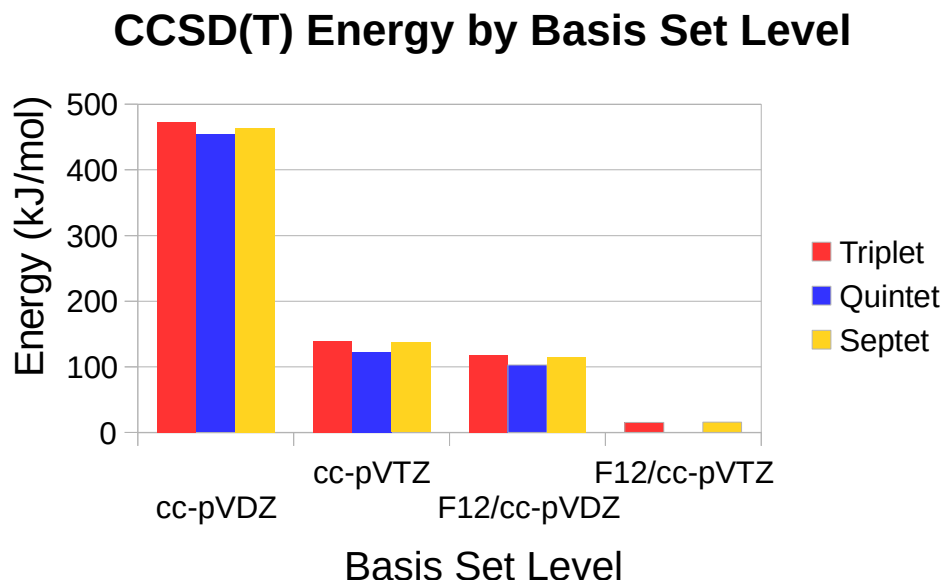


Figure 3.5: RHF/CCSD(T) energies of model **1b** by basis set level, relative to the lowest (RHF/CCSD(T)-F12(cc-pVTZ), $S=2$)

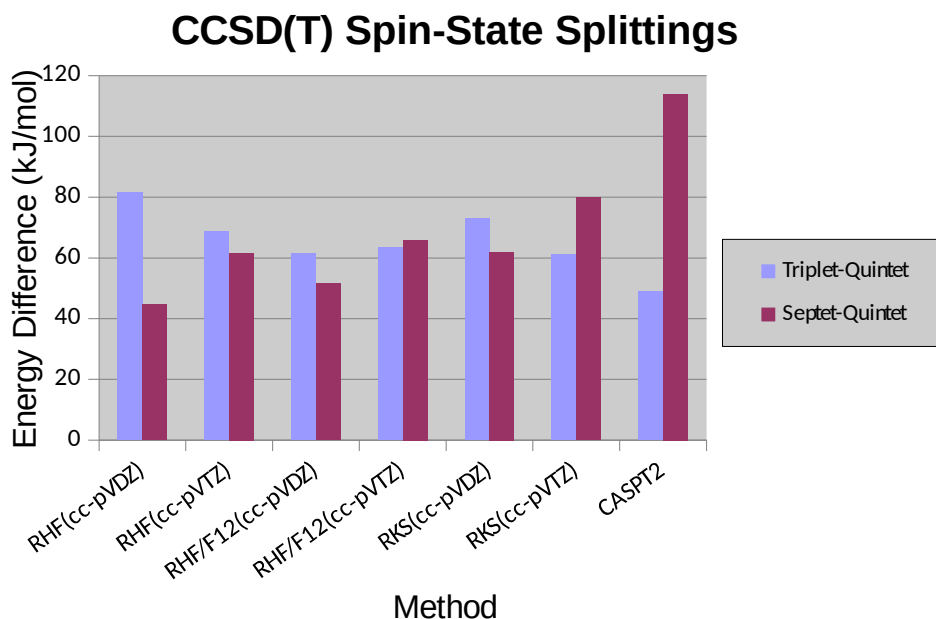


Figure 3.6: Spin-state ordering (relative to quintet) of Model **1b** from CCSD(T) by method

The use of restricted open-shell Kohn-Sham (RKS) reference orbitals, which are often more reliable for multireference systems [34], was also investigated. These generally gave results more closely resembling the CASPT2 results of Borowski *et al.* than did the calculations with RHF reference (Figure 3.6 and Table 3.5). The quintet-triplet sep-

aration was similar, at 49 kJ/mol for CASPT2 and 61 kJ/mol for RKS/CCSD(T)(cc-pVTZ), but the quintet-septet separation was considerably higher for CASPT2 than for RKS/CCSD(T)(cc-pVTZ) (114 kJ/mol for the former but only 80 kJ/mol for the latter). It is unfortunate that attempts to run calculations with both RKS reference orbitals and explicit correlation were unsuccessful, as basis-set convergence does not seem to have occurred at the TZVP level, but convergence for RKS/CCSD(T)-F12 calculations proved problematic, and the large system size makes basis sets above the TZVP level unwieldy.

Whilst the CCSD(T) calculations at higher basis set levels yielded energy separations approaching those obtained from CASPT2, the situation is further complicated by the fact that the system has a highly multireference character (see Section 2.2.1). As such, it is important to consider diagnostics to assess the degree of this multireference character. The T_1 and D_1 diagnostics are shown in Table 3.6. As a general rule, a value of the T_1 much above 0.05 indicates significant static correlation [34]. In a system of perfectly homogeneous multireference character, this cutoff would correspond to a D_1 of 0.07.

Table 3.6: Coupled-cluster diagnostics

Case	Triplet		Quintet		Septet	
	T_1	D_1	T_1	D_1	T_1	D_1
RHF/CCSD(T)(cc-pVDZ)	0.070	0.315	0.032	0.173	0.033	0.147
RHF/CCSD(T)(cc-pVTZ)	0.067	0.300	0.033	0.190	0.032	0.132
RHF/CCSD(T)-F12(cc-pVDZ)	0.071	0.323	0.033	0.159	0.036	0.157
RHF/CCSD(T)-F12(cc-pVTZ)	0.067	0.304	0.033	0.184	0.033	0.136
RKS/CCSD(T)(cc-pVDZ)	0.016	0.059	0.018	0.082	0.018	0.065
RKS/CCSD(T)(cc-pVTZ)	0.016	0.057	0.018	0.080	0.018	0.066

Several trends emerge from the data. Firstly, the diagnostics are largely unaffected by the basis set or the use of F12 explicit correlation. The highest difference in the T_1 diagnostic between the cc-pVDZ and cc-pVTZ basis levels was 0.004, for the case of RHF/CCSD(T)-F12 in the triplet state. Whilst the D_1 diagnostic often shows a higher fluctuation across basis levels in absolute terms, in relative terms it is quite similar. The change in the T_1 diagnostic in the aforementioned case corresponds to a 5.6% decrease, whilst the fluctuation in D_1 diagnostic for the same case, although higher in absolute value at 0.019, corresponds to a 5.9% decrease.

Secondly, the use of Kohn-Sham reference orbitals gave lower values of either diagnostic than Hartree-Fock reference orbitals. The D_1 diagnostic was always reduced by more than 50%, and the T_1 by more than 30%, when a given calculation was run with RKS rather than RHF orbitals.

Thirdly, values of both diagnostics are higher for the triplet than for the quintet or the septet, probably because for the triplet the HOMO-LUMO gap is relatively low (evidenced by the quintet ground state of the complex), reducing the energetic cost of single excitations. This is particularly evident with the use of RHF orbitals. The use of RKS reference orbitals significantly reduces the difference in the values of both diagnostics for the triplet and those for the other two multiplicities, perhaps by calculating more evenly spaced orbitals.

Finally, and most significantly, whilst the T_1 diagnostic had reasonably low values in most cases, the D_1 had significantly higher values. This is likely to be because the orbitals that most significantly contribute to the multireference character are localised to the iron, whilst the ammonia and formate residues brought the average multireference character down. Whilst the values of the D_1 diagnostic with RKS reference orbitals might be considered borderline, even a borderline value of this diagnostic makes these data unreliable as benchmarks.

3.3.4 Spin-State Energetics of Other Models

Due to the uncertain structure of the oxoferryl intermediate of SyrB2, several other isomers of the Model **1b** were considered (Models **2b** and **3b**, see Figure 3.7). Although not of benchmark quality, RKS/CCSD(T) results are still considered alongside those from B3LYP, B3LYP having been chosen for its ability to reproduce experimental and CASPT2 results in the previous sections.

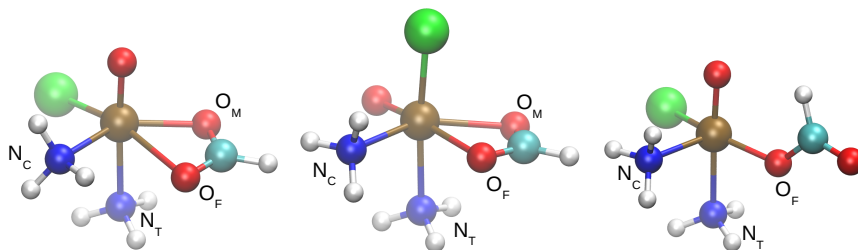


Figure 3.7: From left to right: Models **1b**, **2b** and **3b**

All complexes were optimised at the B3LYP-D/def2-TZVP level. Geometrically, there is very little difference between the corresponding Fe–L bond lengths for first shell ligands for different models (Table 3.7). The trends in bond-length difference between spin states are also borne out in the same manner, with an extension in z-axis Fe=O and Fe-N_T bonds

Table 3.7: Geometric parameters of model complexes, calculated at the B3LYP/def2-TZVP level

	Case	Fe=O	Fe-Cl	Fe-N _C	Fe-N _T	Fe-O _F	Fe-O _M
Triplet	1a	1.62	2.27	1.97	2.18	2.03	2.03
	1b	1.61	2.25	2.02	2.19	2.06	2.03
	2b	1.61	2.24	2.04	2.11	2.30	2.01
	3b	1.59	2.28	1.96	2.15	1.91	n/a
Quintet	1a	1.62	2.30	2.11	2.15	2.36	2.04
	1b	1.61	2.26	2.16	2.16	2.40	2.05
	2b	1.62	2.23	2.22	2.17	2.10	2.24
	3b	1.62	2.25	2.20	2.17	1.87	n/a
Septet	1a	1.89	2.27	2.16	2.27	2.20	2.14
	1b	1.89	2.24	2.21	2.29	2.22	2.14
	2b	1.87	2.24	2.20	2.31	2.31	2.09
	3b	1.90	2.23	2.19	2.30	1.90	n/a

on moving to the septet spin-state. This is particularly interesting, as it is also observed in Model **2b**, in which Fe=O and Fe-N_T do not form a single axis.

Again, the energetic trends of the coupled cluster data and the B3LYP data (Table 3.8) are similar. The quintet is always the ground state, and, where both are available, the triplet is always lower in energy than the septet. Notably, the energy separation between the quintet and both of the excited states is significantly higher in the pentacoordinate species **3b** than in the two hexacoordinate species. Unfortunately, most RKS/CCSD(T) calculations with triplet multiplicity failed to converge. The hexacoordinate species **1b** and **2b** are always lower in energy than the pentacoordinate species **3b**, but the separation is low enough that, in a full protein environment, a five-coordinate species such as **3b** might be preferred due to environmental factors such as a stabilising hydrogen bond.

Table 3.8: Energies of quintet, triplet and septet states of Models **1b**, **2b** and **3b** relative to the lowest energy structure

	Multiplicity	1b (kJ/mol)	2b (kJ/mol)	3b (kJ/mol)
B3LYP	Triplet	48	40	85
	Quintet	0	15	18
	Septet	71	63	106
RKS/CCSD(T)	Triplet	61	—	—
	Quintet	0	13	28
	Septet	80	75	129

3.4 Conclusion

This chapter set out to investigate the electronic structure of the oxoferryl intermediate **D**, and the ability of various computational techniques to represent it, in order to inform subsequent calculations of SyrB2’s mechanism. The ability of DFT and CCSD(T) to reproduce a number of parameters from experiment and from higher level calculations, and the question of which density functional does so most reliably, were investigated.

All functionals tested are able to replicate reasonably the known geometries of the two oxoferryl complexes $\text{trans-[Fe(IV)(O)(TMC)(NCCH}_3\text{)]}^{2+}$ and $[\text{Fe(IV)(O)(TMC-py)}]^{2+}$. In addition, every functional correctly identifies a quintet ground state for models **1a** and **1b**. Of the functionals tested, the one best able to replicate the separation in energy between the quintet state and the triplet and septet states from CASPT2 is B3LYP, a finding that is in accord with the literature. Unfortunately, since any hopes of using CCSD(T) as a benchmark were confounded by the highly multireference character of the system and incomplete basis-level convergence, this study remains reliant on the CASPT2 benchmarks of Borowski and co-workers. Finally, it was observed that models **1b**, **2b** and **3b** are reasonably electronically similar to one another, and that although a hexacoordinate structure is preferred energetically, the preference is sufficiently slight that a pentacoordinate structure could certainly be formed in a full protein environment if stabilising factors such as hydrogen bonds were available.

Chapter 4

The Protein-Substrate Complex

4.1 Introduction

4.1.1 The Need for a Model of the Protein-Substrate Complex

With an improved picture of the electronic structure of SyrB2's Fe(IV) active site complex, this chapter focusses upon the development of a better model of the protein-substrate complex. As discussed in Chapter 1, there is a lot of evidence to suggest that the origins of SyrB2's selectivity lie in substrate positioning, based on the observation that the non-native substrates NVA and ABA react with far less specificity than the native substrate THR [23]. Whilst, structurally, NVA and ABA differ significantly from THR in that they lack its γ -OH group, the reactive C-H bonds of NVA and ABA are chemically essentially identical to that of THR (the dissociation energy for an ethyl C-H bond is $420.5 \text{ kJ mol}^{-1}$, compared to $421.7 \text{ kJ mol}^{-1}$ for those of ethanol's methyl group [92]).

Even in the absence of these findings, however, the pivotal importance of a reliable starting structure to any mechanistic calculations would require little justification. During hydrogen abstraction the position of the substrate is particularly important, as low-lying σ^* and π^* MOs of the Fe=O unit lead to competing but non-degenerate σ - and π - channels to reaction [93], depending on the direction from which the hydrogen approaches (see Section 5.1.2). As a result, the orientation of the C-H bond during its approach to the Fe=O bond can have dramatic consequences to the energy barrier to the reaction.

4.1.2 Insights into Active-Site Structure

Whilst no crystal structure of the protein-substrate complex exists, a large amount of information on their interactions has been derived, both experimentally and computationally. Perhaps most significantly, Fullone and co-workers [16], through a combination of a

docking study and a mutational analysis, identified a number of residues in the active site whose mutation to alanine hindered or prevented a reaction with SyrB2. A 2-dimensional representation of these residues is shown in Figure 4.1.

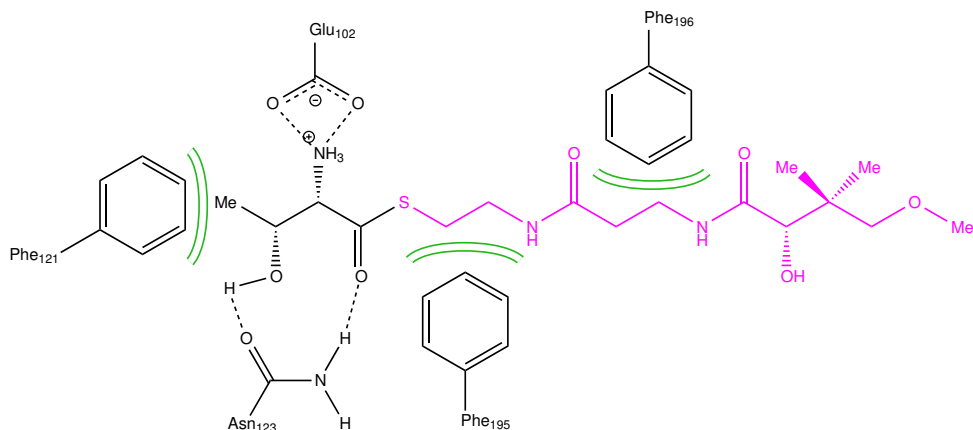


Figure 4.1: Protein-substrate interactions

The two phenylalanine residues Phe121 and Phe195 were proposed to offer hydrophobic stabilisation to the substrate head and tail respectively, whilst Asn123 was proposed to hydrogen-bond to the more polar groups of the substrate head. Glu102 was found to form a salt bridge to the substrate amino acid group by both Fullone and co-workers and a prior docking study [31]. Interestingly, although important to reactivity, Phe196 blocks access to the active-site pocket in the crystal structures. Previous docking studies [94, 31] have rotated the side chain of this residue prior to docking.

Although not included in this mutational analysis, another nearby residue, Arg254, is also believed to be significant to reactivity. It has been postulated variously to form a hydrogen bond to the ferric hydroxide group of **E** (deactivating it during radical rebound) [29], a network involving Glu102 and succinate in a hexacoordinate iron complex [31], and succinate in a pentacoordinate iron complex, stabilising the monodentate succinate [26].

In spite of this wealth of study of protein-substrate interactions, to date there has been very little study of the docking of non-native substrates to SyrB2, and no MD simulations of SyrB2, with or without any docked substrates. In this chapter, the intermolecular interactions between SyrB2 and the three substrates THR, NVA and ABA are studied, through a comparative docking study of these three substrates, and through MD simulations, with these three substrates and with both Fe(II) and Fe(IV) active site complexes. These simulations show THR, through interactions of its γ -OH group, to adopt a different conformation to the other two substrates, leading both to a different position of its reactive carbon relative to the active-site iron complex, and the disruption of an important hydrogen-bonding network in the active-site channel.

4.2 Computational Details

4.2.1 Docking

As 2FCT, the PDB structure of SyrB2 with bound Fe, Cl and 2OG [12], has several missing residues (Met1-Ser2 and Ile57-Ser58-Gly59-Gly60), a completed protein structure, taken from the MSci Thesis of Andrew Jarnuczak [95], was used. This structure was prepared using the program Modeller [96], which creates potential structures for missing sections by constructing an initial loop model, randomly displacing it to generate a number of structures, then carrying out an optimisation of each of these structures at the MM level. 100 such optimised structures were generated, and the most favourable selected according to its DOPE-HR [97] score and its RMSD from 2FCT. Next, hydrogen atoms were added using Reduce [98] automated by the MolProbity webserver [99]. The residues Gln11, Asn95, Gln129 and Gln245 were all flipped from their conformation in the crystal structure. Finally, the protonation states of all titratable residues were calculated using PropKa [100]. Protonation states of histidine residues are shown in Table 4.1. All other residues were left in their standard protonation states.

Table 4.1: Protonation states of histidine residues

Residue	Protonation State
His69	N $^{\epsilon}$
His78	N $^{\epsilon}$
His116	N $^{\delta}$
His235	N $^{\delta}$
His240	N $^{\epsilon}$
His261	N $^{\delta}$
His268	N $^{\epsilon}$
His300	N $^{\epsilon}$

Docking was carried out using Autodock VINA 1.1.2 [68], as the substrate is quite large and VINA has been shown to be capable of reliably handling large numbers of rotatable bonds [67, 68]. During the docking of all three substrates, the substrate was attached to a methyl pantetheinyl ether arm to represent the pantetheinyl carrier section of SyrB1 (see Figure 4.1). The entire substrate-pantetheine complex was treated as flexible, with the exception of amide bonds, thioester bonds and bonds whose rotation would not affect molecular symmetry. Three protein bonds were also designated as rotatable. These were the C-OH bond of Tyr272, whose movement could potentially allow hydrogen bonding (see Figure 4.2), and the C $^{\alpha}$ -C $^{\beta}$ and C $^{\beta}$ -C $^{\gamma}$ bonds of Phe196, which in the crystal structure

blocks access to the active site, so must necessarily move to clear a path for the ligand. All docking runs used the same box, of $24 \times 20 \times 16$ Å, and produced 20 poses. Default settings were used for all other parameters.

4.2.2 Molecular Dynamics with Apoprotein or Fe(II) Active Site Complex

Molecular dynamics commenced from the most favourable docked structure for each substrate. In addition, simulations of the holoprotein and apoprotein (using the same completed protein structure used for docking) were run. Explicit solvent (TIP3P water [101]) was added, followed by sodium ions to neutralise the charge. During all simulations (with the exception of that of the apoprotein), iron and first coordination shell ligands were held in place with position restraints of 10^4 kJ mol⁻¹ nm⁻¹.

The Gromacs 4.5.6 suite of programs [102] was used throughout. The topology was generated using `pdb2gmh`, using the Amber ff03 forcefield [103] for the protein. GAFF parameters with Merz-Singh-Kollman [104] (MSK) charges were used for the substrate and the iron complexes, whose topologies were generated using Antechamber [105] implemented through ACPYPE [106]. Charges for the iron complex and the substrate THR came from the MSci thesis of Andrew Jarnuczak [95]. For the cofactors, charges were calculated using Turbomole at singlet multiplicity at the TPSS/def2-TZVP/cosmo($\epsilon = 80$) level, following an optimisation under the same conditions. Charges for THR were calculated at the B97-D/def2-TZVP level following an optimisation at the same level. Charges for NVA and ABA were calculated using Gaussian09 at the B97D/def2-TZVP/pcm level with water solvent, following an optimisation at the same level, with a distance constraint from the beginning to the end of the pantetheine arm set to the same distance as for THR, to ensure the same conformation.

After preliminary energy minimisation and pressure equilibration (Table 4.2), the simulations were run in an NVT ensemble until equilibrated, then for a further 17–27.5 ns. In all cases the leap frog integrator was used, with a step size of 2 fs. Simulations run in an NPT ensemble made use of the Berendsen barostat [66], whilst those run in an NVT ensemble made use of the v-rescale thermostat [65]. All bonds were constrained. Periodic boundary conditions were employed, using the PME method for long range electrostatics. Van der Waals cut-offs and Coulomb cut-offs were both set to 1 nm.

Table 4.2: Equilibration procedure for MD simulations with Fe(II) active-site complex

Parameter	Energy Minimisation	Temperature Equilibration	Equilibration Stage 1	Equilibration Stage 2	Equilibration Stage 3	Final Equilibration/Production
Ensemble	-	NVT	NPT	NPT	NPT	NVT
Integrator	Steepest descent			Leap-frog		
Duration (ps)	-	100	400	100	400	Until equilibration
Constraints	All bonds to hydrogen			All bonds		
Position Restraints (kJ mol ⁻¹ nm ⁻¹)	1000	1000	1000	100	0	
τ_T	-	0.1	0.1	0.1	0.1	2
τ_P	-	-	1	1	1	-

4.2.3 Molecular Dynamics with Fe(IV) Active-Site Complex

Next, simulations were run with the holoprotein with all fifteen combinations of the five Fe(IV) active site complex isomers shown in Figure 1.7 and the three substrates THR, ABA and NVA. To prepare the simulations with Fe(IV) active-site complexes, a representative equilibrated snapshot was taken from each of the simulations with Fe(II) active-site complexes with bound substrate. The coordinates of the Fe complex (iron and all first coordination sphere ligands) were taken, modified manually to the Fe(IV) active-site complexes and optimised at the B3LYP/def2-SVP level, then replaced into the protein. The succinate and substrate were then optimised at the MM level with the rest of the system frozen, to prevent steric clashes.

MSK charges were calculated for these model complexes at the B3LYP/def2-TZVP+/pcm level with 1-fluorooctane solvent. Again, all other parameters for these complexes came from GAFF, with topologies prepared using Antechamber implemented through ACPYPE.

The simulations were subjected to a 100 ps temperature equilibration phase (as in Table 4.2), before a production run of 20 ps.

4.2.4 Analysis

In this study, several metrics are used in the analysis of results. The root-mean-square deviation (RMSD) is a useful measure of the difference between one structure and another [32]. This can be used for simple comparison of two structures or for comparison of each frame of an MD trajectory to a starting structure, to show the magnitude of conformational changes.

$$\text{RMSD} = \sqrt{\frac{1}{M} \sum_{i=1}^N m_i ||(x_i(t) - x_i(ref))||^2} \quad (4.1)$$

Here N refers to the total number of atoms, m_i to the mass m of a given atom i , M to mass of the entire system, and $x_i(t)$ and $x_i(ref)$ to the atomic coordinates of atom i in structure t and the reference structure.

Root-mean-square fluctuation (RMSF) is a measure of the level of fluctuation of a structure from a reference position over a given set of timesteps. Although calculated on an atom-by-atom basis, it is useful for quantifying the mobility of residues or entire proteins.

$$\text{RMSF} = \sqrt{\frac{1}{T} \sum_{t=1}^T (x_t - x_{ref})^2} \quad (4.2)$$

Here t refers to a given timestep, T to the total number of timesteps, x_t to the atomic positions at timestep t and x_{ref} to the reference structure.

Molecular dynamics trajectories were tested for equilibration by running a series of tests on the backbone RMSD (after binning), following the procedure of Senn *et al.* [107, 108]: The Mann-Kendall test for trend in bin means and standard deviations, the Shapiro-Wilk test for normality of bin means, and the von Neumann test for correlation in bin means.

Caver 3.0 [109] was used to search molecular dynamics trajectories for channels between the active site chamber and the bulk. This treats a protein structure as a set of hard spheres of fixed radius on a grid, identifying a surface by means of a shell probe. Tunnels can then be identified on the grid between a given node inside the hull and the nearest reachable node outside the hull.

Trajectories are searched for tunnels frame by frame. Tunnels in each frame j are assigned a cost (Equation 4.3), depending on their length (measured by the total number of nodes i that they contain, N_n), and the maximum radius of empty space surrounding each node (r). Similar tunnels from different frames are grouped together into channels, which are assigned priorities (Equation 4.4) according to their cost, C , and the number of frames, N_f , in which they appear.

$$C = \frac{1}{N_n} \sum_{i=1}^{N_n} \frac{1}{r_i^2} \quad (4.3)$$

$$\text{Priority} = \frac{\sum_{j=1}^{N_f} e^{-C_j}}{N_f} \quad (4.4)$$

In each case, the central iron was selected as the starting point for any tunnels, and a probe radius of 0.9 Å was used. A shell radius of 3 Å and shell depth of 4 Å were used to define the protein surface.

4.3 Results and discussion

4.3.1 Docking

Each substrate was docked to the completed protein structure, with twenty conformations considered per substrate. For THR, this yielded four conformations of the twenty in which the reactive carbon, C γ , was close enough to the bound chloride for a reaction to commence without a significant conformational change. These four conformations also had the highest binding affinity. In all four, Phe196 rotated out of the way in order to yield access to the active site, giving the same conformation in each case. In none of these four conformations was Tyr272 able to interact significantly with the substrate amino or γ -OH groups, which instead generally chose to form hydrogen bonds or salt bridges to one or both of the nearby residues Arg254 and Glu102. These four poses are shown in Figure 4.3, with the affinities and RMSDs of all 20 conformations shown in Table 4.3.

The four conformations with highest binding affinity were structurally very similar—sufficiently so that a molecular dynamics simulation of any reasonable length would be expected to explore their conformational space. In all of them, Phe195 and Phe196 provided hydrophobic stabilisation in the entrance to the substrate channel **T**₁, in accord with the findings of the mutational analysis study of Fullone *et al.* [16]. Poses 1–3 (panels **A** to **C** of Figure 4.3) differ primarily in the hydrogen bonding partner of the γ -OH group, (either Arg254 or Glu102). The most different pose structurally was Pose 4 (panel **D** of Figure 4.3), which was the only one not to maintain a salt bridge between the substrate amino group and Glu102, an interaction identified as particularly important by previous docking studies [31, 16], as well as by the mutational analysis of Fullone *et al.* Pose 4 also ignores a number of other potentially stabilising interactions (the substrate’s amino group and hydroxyl groups are both without a hydrogen bonding partner), and has a proportionately lower binding affinity. The most energetically favourable conformation (Figure 4.2) was chosen to proceed to MD simulations.

NVA and ABA each also gave a small set of chemically reasonable poses from the twenty generated. Again, from their similarity to one another one might reasonably expect them to be explored from one another over the course of an MD simulation (see Figure 4.4 and Table 4.4). Interestingly, although the different reactivity strongly suggests that THR should have a different placement to the other two substrates, the most favourable pose

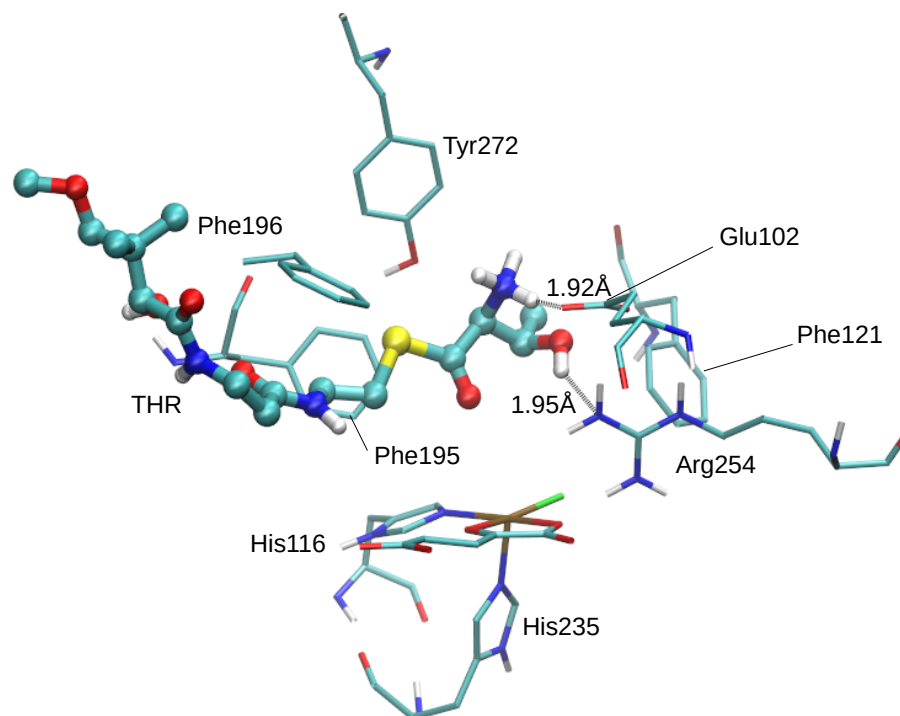


Figure 4.2: Most favourable docked conformation of THR

for each substrate is almost identical.

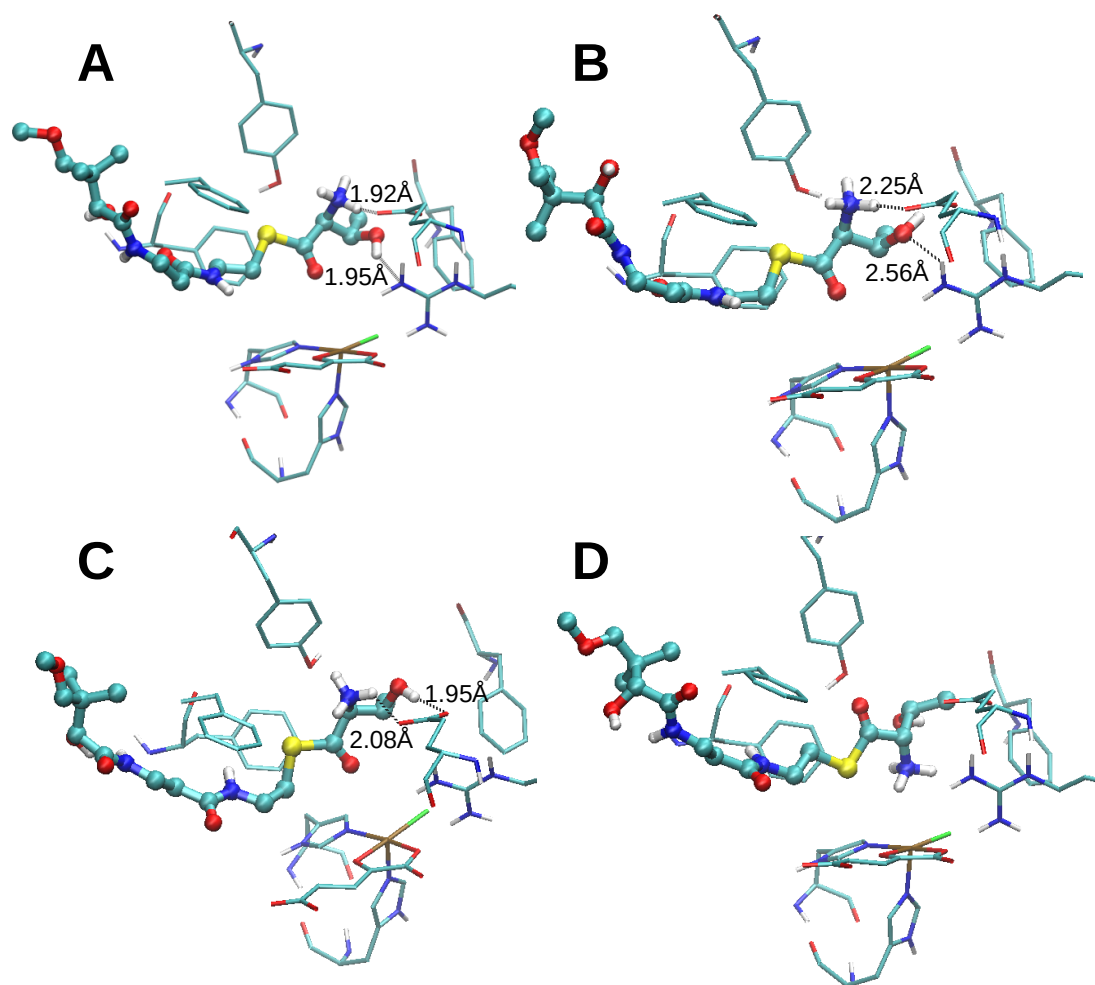


Figure 4.3: **A–D**: Four most favourable binding conformations of THR, in descending order

Table 4.3: Docked poses of THR

Pose	Affinity (kJ mol ⁻¹)	RMSD (Å)
1	-33.5	0.00
2	-32.6	1.41
3	-32.2	1.71
4	-31.4	1.68
5	-29.7	9.96
6	-29.7	3.29
7	-29.3	9.71
8	-28.9	9.36
9	-28.9	9.70
10	-28.9	10.13
11	-28.9	10.17
12	-28.9	10.57
13	-28.5	8.63
14	-28.5	9.69
15	-28.5	9.99
16	-28.5	9.81
17	-28.5	9.82
18	-28.5	10.36
19	-28.0	10.45
20	-28.0	10.43

Table 4.4: The ten docked poses of ABA and NVA with highest affinity

Pose	ABA		NVA	
	Affinity (kJ mol ⁻¹)	RMSD (Å)	Affinity (kJ mol ⁻¹)	RMSD (Å)
1	-32.6	0.00	-31.8	0.00
2	-32.2	1.88	-31.4	1.69
3	-31.4	2.08	-31.0	1.86
4	-30.1	2.44	-30.1	2.26
5	-29.7	9.33	-28.9	2.06
6	-29.3	5.19	-28.5	9.75
7	-29.3	9.51	-28.0	9.60
8	-29.3	9.59	-28.0	5.46
9	-29.3	9.67	-28.0	9.43
10	-29.3	9.70	-28.0	7.86

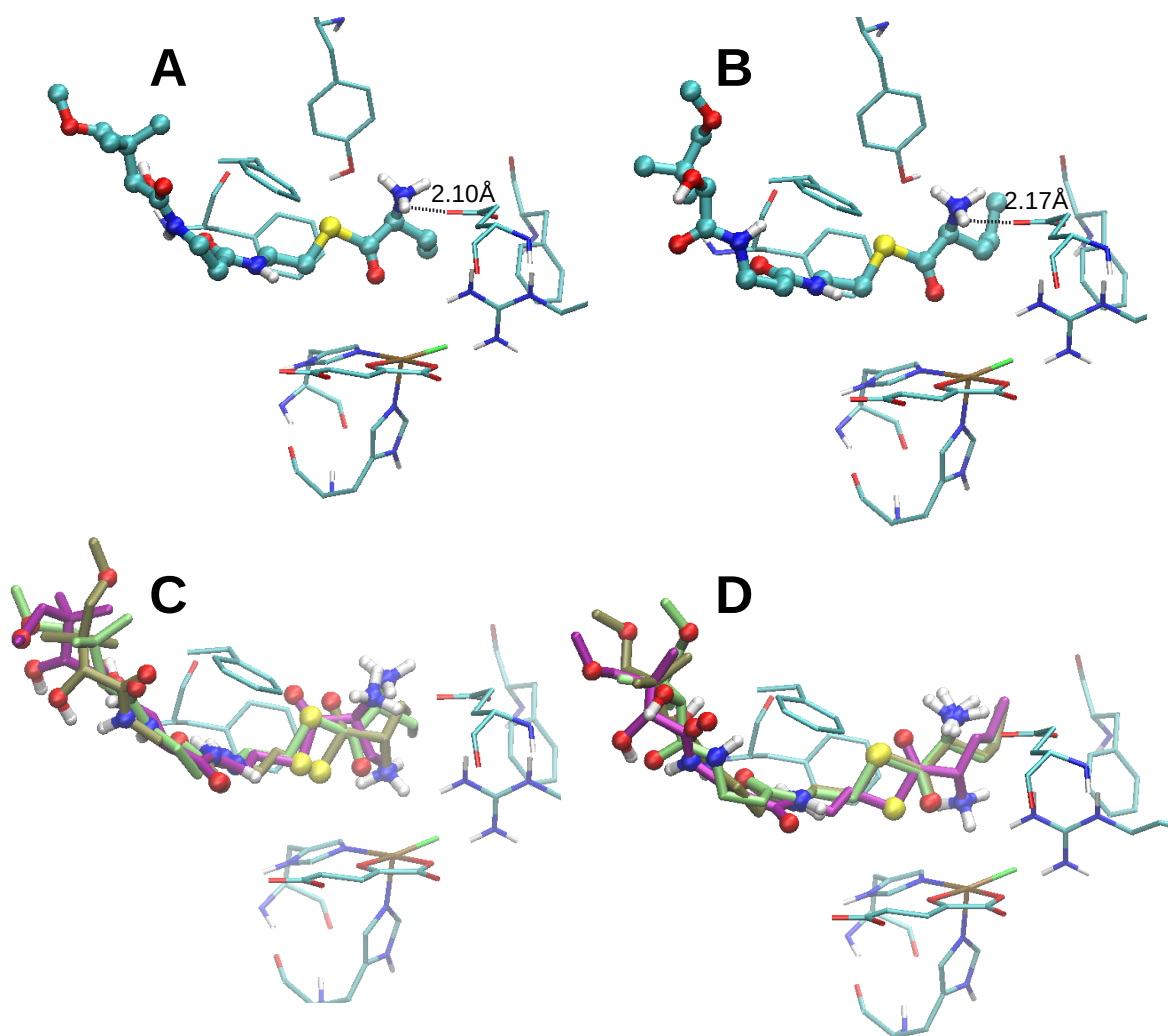


Figure 4.4: **A** and **B**: Most favourable docked conformations of ABA and NVA respectively. **C** and **D**: Overlay of 3 most favourable poses of ABA and NVA respectively

4.3.2 Molecular Dynamics with Iron (II) Active-Site Complex

Substrate Placement and Protein Mobility

Initially, five simulations were prepared—the holoprotein, the apoprotein, and the holoprotein with each of the three substrates bound. RMSDs are shown in Table 4.5. The most mobile residues were the same in each case—generally the two modelled sections (the Met1-Ser2 head and the Ile57-Ser58-Gly59-Gly60 loop), the Val310 tail and several surface loops (Table 4.7).

Table 4.5: Backbone RMSD from crystal structure

Simulation	RMSD(Å)	Standard Deviation (Å)
Apoprotein	1.24	0.10
Holo	1.15	0.09
THR	1.95	0.12
ABA	1.33	0.08
NVA	1.24	0.08

Table 4.6: Substrate–protein interactions. d_{OH} refers to the minimum distance between either carboxylate oxygen and any of the substrate ammonium hydrogens. d_{RH} refers to the minimum distance between any atom of the phenylalanine ring to any atom of the substrate. d_{CCl} refers to the distance between the reacting carbon and the chloride.

Substrate	Structure	Glu102	Phe121	Phe195	Phe196	d_{CCl} (Å)
		d_{OH} (Å)	d_{RH} (Å)	d_{RH} (Å)	d_{RH} (Å)	
THR	Docked Pose 1	1.92	3.64	3.46	3.36	5.24
	MD Average	3.85	5.44	3.23	2.86	4.07
NVA	Docked Pose 1	2.17	3.75	3.49	2.16	6.10 (C^δ)
	MD Average	1.74	4.42	2.85	2.95	4.27 (C^δ)
ABA	Docked Pose 1	2.10	3.73	3.52	3.10	4.02
	MD Average	1.82	5.25	3.10	2.74	4.21

ABA and NVA maintained their docked conformations, with most of the interactions from docking left intact (see Table 4.6). THR, on the other hand, adopted an entirely new conformation: during equilibration, its γ -OH group broke its hydrogen bond to Arg254 and, after short-lived interactions with the iron-bound chloride and the carboxylate group of Glu102, formed a stable hydrogen bond to the tail of 2OG (Figure 4.5). This had implications to both the position of THR and the elaborate hydrogen-bonding network

Table 4.7: RMSFs of flexible residues. Modelled sections were the Met1-Ser2 head and the Ile57-Ser58-Gly59-Gly60 loop. Glu137-Phe138 were part of a particularly mobile surface loop, and Val310 was the tail terminus

Residue	Apoprotein RMSF(Å)	Holoprotein RMSF(Å)	THR RMSF(Å)	NVA RMSF(Å)	ABA RMSF(Å)
Met1	6.4	4.9	2.3	6.6	4.4
Ser2	5.2	3.6	1.2	4.1	2.4
Lys3	2.5	1.7	2.4	1.9	1.2
Ile57	3.0	2.7	2.7	2.2	3.3
Ser58	2.9	2.2	2.2	3.0	2.9
Gly59	2.8	2.2	1.5	2.7	2.5
Gly60	2.2	2.0	0.7	1.5	1.4
Glu137	2.7	2.4	2.0	2.5	3.1
Phe138	2.5	2.2	2.2	2.8	2.5
Val310	3.8	2.5	3.5	2.5	2.7
No. other residues with RMSF ≥ 2.3 Å	6	3	5	4	5

surrounding the tail of 2OG, which in other simulations was reasonably stable (Figure 4.7).

As a result of this interaction, THR was able to pull further out of \mathbf{T}_1 than the other two substrates (Figure 4.6). This resulted in the cleaving of the direct threonyl ammonium-Glu102 salt bridge, which was replaced with a water-mediated interaction. However, due to a different conformation of the $\text{N} - \text{C}^\alpha - \text{C}^\beta - \text{C}^\gamma$ torsion, this did not result in a greater distance between the reacting carbon and the chloride.

Accessibility of the active site channel and conformation of “gatekeeper” Phe196

As discussed in Section 4.2.1, the substrate channel can be opened or closed by the side chain of Phe196, primarily by rotation around the $\text{C} - \text{C}^\alpha - \text{C}^\beta - \text{C}^\gamma$ torsion (Figure 4.9). In the crystal structure 2FCT, this had a value of -166.6° . The docked structures all had an almost identical conformation of this angle (135° , 133° and 134° for THR, NVA and ABA respectively). During molecular dynamics simulations in the presence of docked substrate, however, the torsion generally had averages at around 60° or 180° (Figure 4.8). During the MD trajectory of the holoprotein, the phenylalanine predominantly remains in the same conformation as in the crystal structure.

Interestingly, although in all simulations the substrate channel was at least partly obstructed, either by the substrate or by the side chain of Phe196, it does not remain

Hydrogen Bonding Partners of the THR OH group

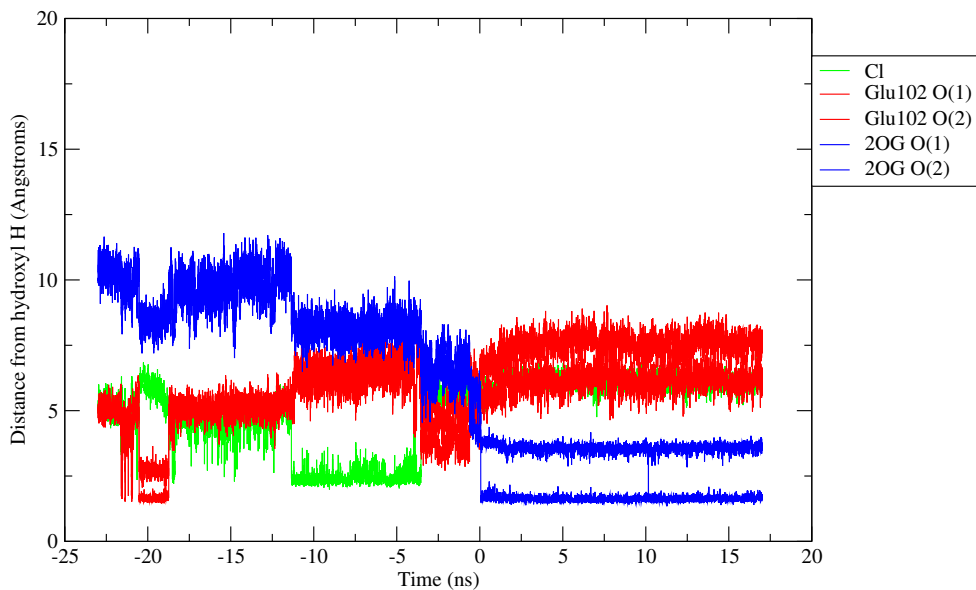


Figure 4.5: Hydrogen bonding partners of the THR γ -OH group. 2OG O(1) and (2) refer to the two oxygen atoms of the oxoglutarate tail carboxyl group, Glu102 O(1) and (2) to the two oxygens of the Glu102 carboxyl group. The equilibrated time stretch begins at 0 ns.

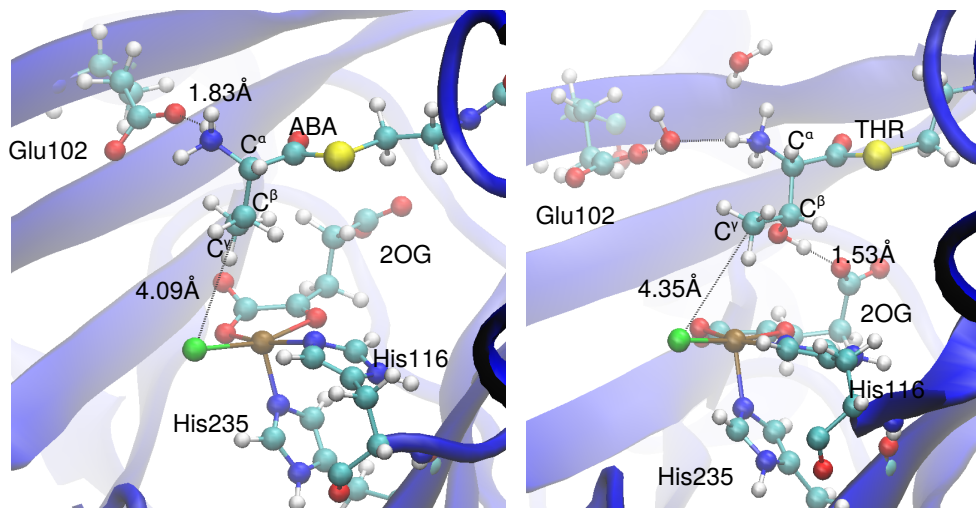


Figure 4.6: Position of ABA (left) and THR (right) following MD equilibration. Note the hydrogen bond between THR's hydroxyl group and 2OG. Note the different states of the $N - C^{\alpha} - C^{\beta} - C^{\gamma}$ torsion.

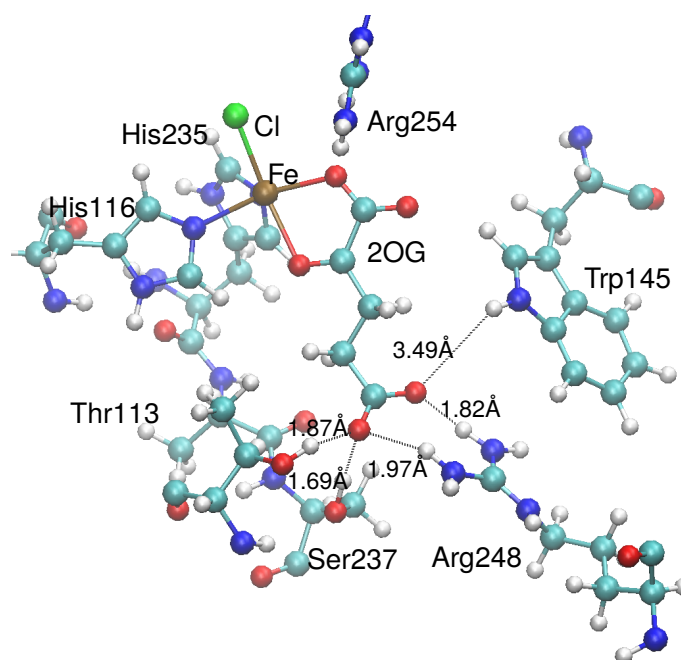


Figure 4.7: 2OG hydrogen bonding network (from simulation with bound NVA, not shown)

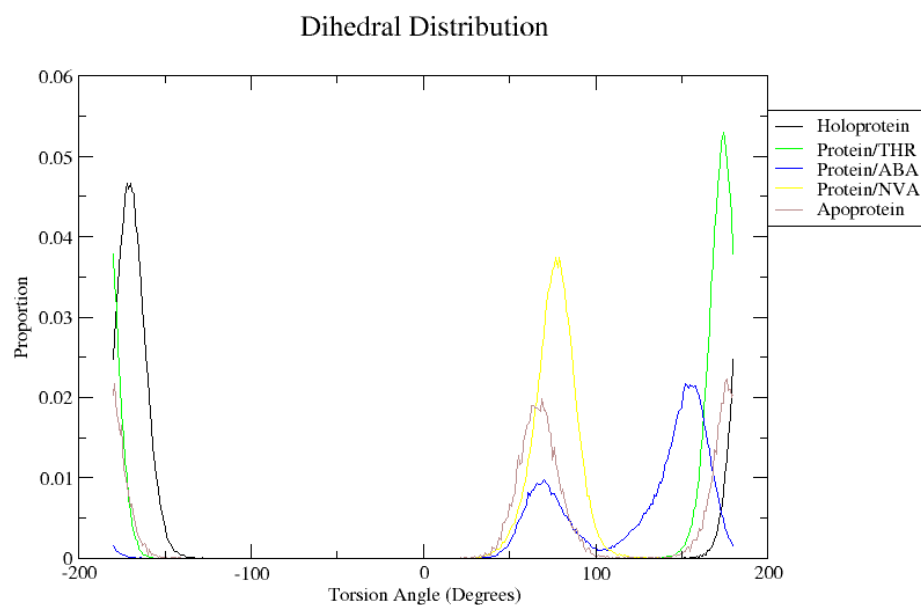


Figure 4.8: Distribution of the $C - C^{\alpha} - C^{\beta} - C^{\gamma}$ torsion of Phe196

entirely closed—a feature that has implications to the reaction cycle, as several small molecules (O_2 and CO_2) are produced or consumed during the reaction. To test the accessibility of the active-site chamber to small molecules, the program Caver was used to

search for tunnels to the active site chamber.

Although neither of the tunnels had a high priority in any simulation, both were observed in every simulation (Table 4.8). Surprisingly, the presence of bound THR increased the throughput of both tunnels in comparison to the holoprotein, despite the obvious steric obstruction caused by THR within \mathbf{T}_1 . This was probably due to the conformation of Phe196, which was held wider open in the simulation with THR than in the simulation with holoprotein only. THR had a particularly open \mathbf{T}_1 compared to the other substrates, as the Glu102-substrate ammonium salt bridge was water mediated rather than direct, meaning both that the substrate was further out of the channel (leaving it wider open), and that the bridge itself did not obstruct the passage of solvent (Figure 4.10).

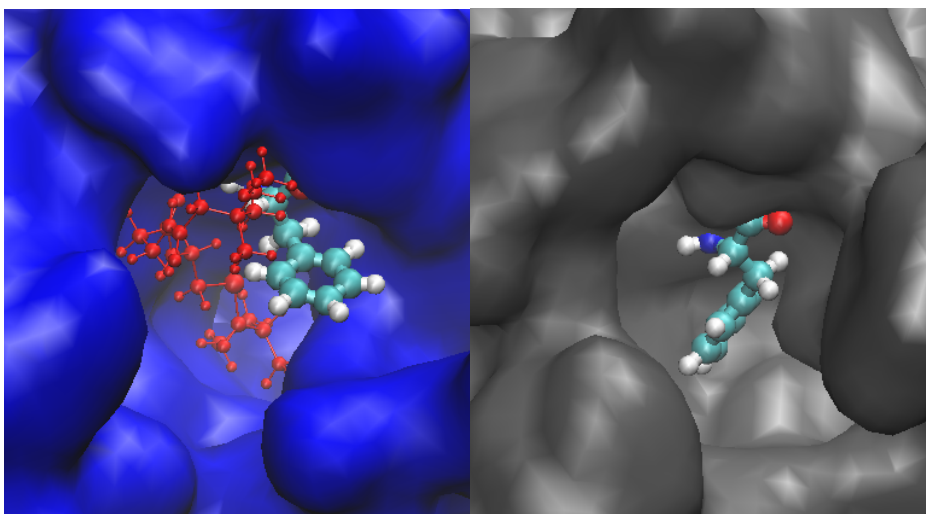


Figure 4.9: Phe196 in protein with bound ABA (left, $C - C^\alpha - C^\beta - C^\gamma = 65.75^\circ$, substrate shown in red) and holoprotein (right, $C - C^\alpha - C^\beta - C^\gamma = -163.05^\circ$)

Table 4.8: Throughputs of channels \mathbf{T}_1 and \mathbf{T}_2 . For the structure with bound ABA, channels were found both above and below the substrate within \mathbf{T}_1

Simulation	\mathbf{T}_1			\mathbf{T}_2		
	Priority	Bottleneck Radius (Å)	Average Length (Å)	Priority	Bottleneck Radius (Å)	Average Length (Å)
Holoprotein	0.192	0.96	21.7	0.081	0.96	23.6
THR	0.280	1.00	16.1	0.203	1.00	22.3
NVA	0.008	0.96	15.7	0.028	0.94	22.7
ABA(1)	0.031	0.96	14.8	0.008	0.96	25.1
ABA(2)	0.009	0.93	29.1			

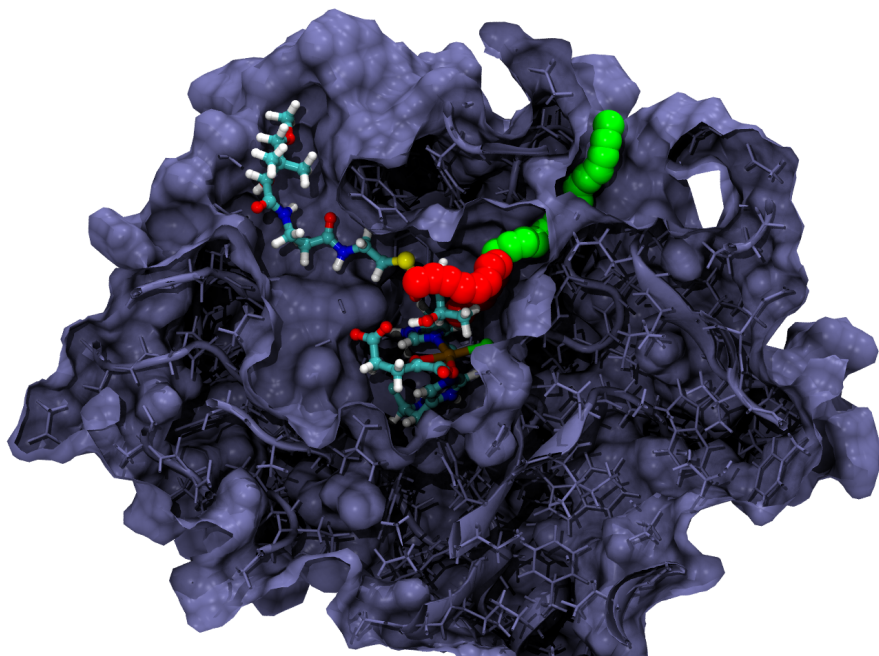


Figure 4.10: Substrate channel **T₁** (red) and allosteric channel **T₂** (green), from Caver analysis of the trajectory of protein with bound THR (shown)

4.3.3 Molecular Dynamics with Iron (IV) Active-Site Complex

The position of the substrate within **T₁** is described relative to other atoms or residues. Two parameters that are particularly useful are the distance between the oxoferryl oxygen and the nearest abstractable hydrogen, d_{OH} , and the distance between the reacting carbon and the chloride ion, d_{CCl} (Figure 4.11). As d_{OH} and d_{CCl} correspond to the reactive atoms in the hydrogen abstraction and chlorine rebound mechanistic steps respectively, these distances have chemical as well as geometric significance. For NVA, which has two reactive carbons, d_{OH} and d_{CCl} are considered for both.

THR

The position of the THR substrate was dependent in each case on the interactions of the γ -OH group. In the presence of the two active-site complexes in which the oxoferryl oxygen was *trans* to His235 (**D2** and **D3**), the threonyl γ -OH group formed a hydrogen bond to it, leading to low values of d_{OH} (with averages of less than 3 Å) but high values of d_{CCl} , in both cases in excess of 6 Å (Figure 4.12, left panel, and Table 4.9). In the presence of the two active-site complexes in which the chloride ion was *trans* to His235 (**D4** and **D5**), the threonyl γ -OH group maintained its hydrogen bond to the succinate tail, leading to a high value of both distances—in both cases d_{OH} was in excess of 5 Å,

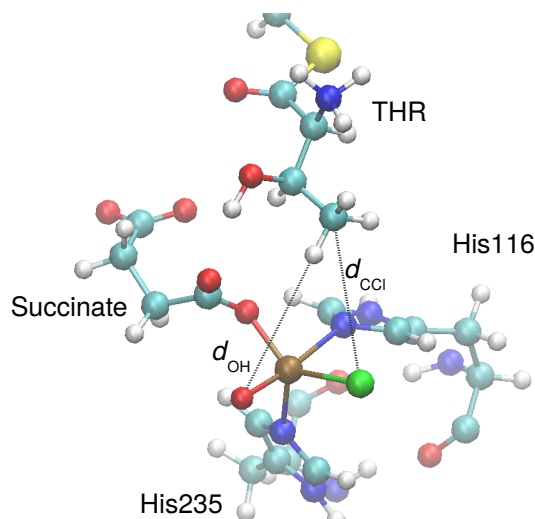


Figure 4.11: Distances d_{OH} and d_{Cl} between THR and the active-site complex isomer **D1**

whilst d_{Cl} was also extremely high (Figure 4.12, right panel). Finally, in the presence of **D1**, the threonyl γ -OH group alternated between the succinate’s head and tail as hydrogen bonding partners (see Figure 4.11 and Figure 5.16). This led to relatively low values of both d_{OH} and d_{Cl} , with the average value of both distances under 5 Å. This makes both reactions feasible, which is in accord with the proposals of Wong and co-workers [26], who suggest this to be the reactive isomer for this substrate.

Table 4.9: Placement of THR relative to oxoferryl complexes

Isomer	THR-OH H-bond	Average d_{OH} (Å)	Average d_{Cl} (Å)
D1	Succinate head	4.17	4.36
	Succinate tail	4.41	4.50
D2	Fe=O	2.79	5.36
D3	Fe=O	2.85	6.12
D4	Succinate tail	5.87	4.00
D5	Succinate tail	7.43	6.42

ABA

ABA, in all but one simulation, did not interact directly with the iron complex, maintaining a conformation very similar to its docked structure. In spite of this lack of direct interaction, in most simulations it maintained low values of d_{OH} and d_{Cl} , below 5 Å in each case (see Table 4.10). The only exception was the simulation with the isomer

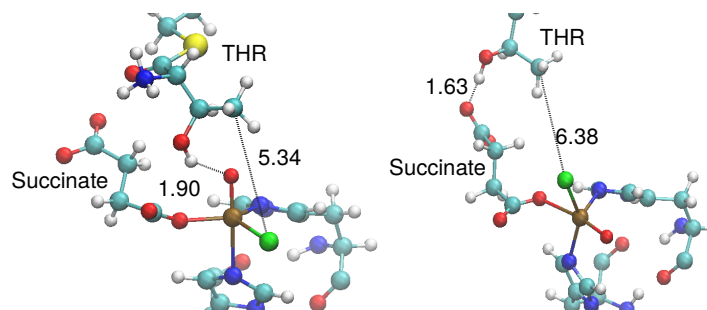


Figure 4.12: Position of THR relative to the **D2** (left) and **D5** (right) active-site complex isomers

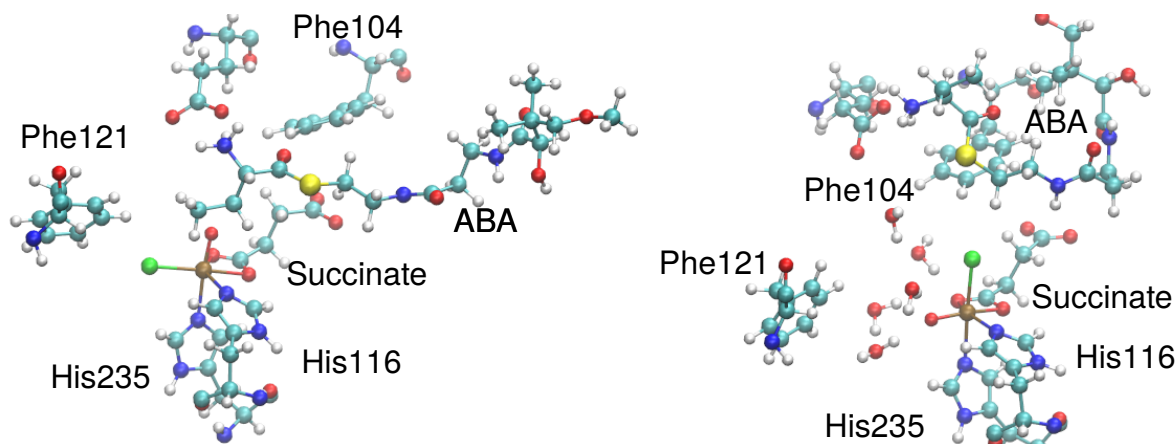


Figure 4.13: ABA in a stable conformation next to the active-site complex isomer **D3** (left) and in an unstable conformation next to the isomer **D4** (right)

D4, in which ABA broke free of the hydrophobic pocket towards the substrate head, and remained in an unstable conformation for the remainder of the simulation (Figure 4.13).

NVA

NVA largely also maintained its docked conformation, again having little or no direct interaction with the iron complex. The situation was somewhat complicated by the presence of multiple conformations of the $C^\beta-C^\gamma$ torsion. Most simulations yielded two stable conformations of this torsion, typically one at around 170° and another at around 270° . Using a cutoff of 5 Å for a plausible reaction, for all active-site complex isomers with the exceptions of **D1** and **D2** at least one of distances d_{OH} and d_{CCl} for one of the two reactive carbons was prohibitively high in both states (Table 4.11). The value of $C^\gamma d_{CCl}$ was prohibitively high for **D3**, and $C^\gamma d_{OH}$ for both states of both **D4** and **D5**. Allowing for the arbitrary nature of the 5 Å cutoff, these results are again in accord with the findings of Wong and co-workers [26], which suggest that **D2** is the reactive species for this substrate.

Table 4.10: Placement of ABA relative to oxoferryl complexes

Isomer	Average $d_{\text{OH}}(\text{\AA})$	Average $d_{\text{CCl}}(\text{\AA})$
D1	4.59	4.22
D2	3.06	4.56
D3	3.15	4.34
D4	No stable configuration	No stable configuration
D5	4.66	4.06

Table 4.11: Placement of NVA relative to oxoferryl complexes

Isomer	$\text{C}^\beta\text{--C}^\gamma$ Torsion	Occupancy (%)	Average C^δ $d_{\text{OH}}(\text{\AA})$	Average C^δ $d_{\text{CCl}}(\text{\AA})$	Average C^γ $d_{\text{OH}}(\text{\AA})$	Average C^γ $d_{\text{CCl}}(\text{\AA})$
D1	250°-280°	66	4.18	4.31	5.26	4.09
	80°-110°	4	3.74	3.98	4.51	4.04
D2	140°-190°	84	3.06	4.29	3.25	5.23
	60°-90°	10	4.85	5.17	2.68	4.39
D3	140°-195°	74	3.06	4.26	3.50	5.29
D4	145°-200°	76	4.70	4.10	6.11	4.25
	270°-320°	18	6.02	4.64	5.63	4.23
D5	240°-270°	52	5.09	4.07	5.11	3.92
	150°-200°	34	4.57	4.14	6.12	4.29

4.3.4 Hydrogen Bonding Environment of the Oxoferryl Species

The hydrogen-bonding environment of the oxoferryl intermediate is also a property of interest, as the presence of hydrogen bonds to water can have mechanistic implications. Indeed, Wong and co-workers [26] have proposed that the presence of multiple oxoferryl species in the Mössbauer spectrum are the result of different hydrogen bonding environments of the $\text{Fe}=\text{O}$ unit. To investigate this proposal, the trajectories of the isomers **D1** and **D2** were searched for the hydrogen-bonding environments of the oxoferryl units. The results of this analysis are shown in Table 4.12.

Experimentally, the relative proportions of the two Fe(IV) species are 4:1 in THR and 7:1 in ABA [27]. Here, the ratios observed are $\approx 1.5:1$ for 0 vs 1 hydrogen bonds in THR/**D1** or $\approx 2.5:1$ for 1 vs 2 hydrogen bonds in THR/**D2**, neither of which is particularly close to the 4:1 ratio of oxoferryl species from the Mössbauer data. In ABA, however, a ratio of 8:1 for 0 vs 1 H-bonds in ABA/**D2** is observed. As such, the results for ABA are in reasonable agreement with the proposal of Wong and co-workers, whilst those for THR

Table 4.12: Hydrogen-bonding environments of oxoferryl species

Substrate	Isomer	% of frames with 1 H-bonds to Fe=O	% of frames with 2 H-bonds to Fe=O	% of frames with 3 H-bonds to Fe=O
ABA	D2	88	11	0
	D1	33	52	14
NVA	D2	33	52	14
	D1	26	68	6
THR	D2	26	68	6
	D1	38	55	7

are not.

4.4 Conclusions

The three substrates have an almost identical docked conformation, which features a direct salt bridge between the substrate ammonium and Glu102. This is in accord with the docking studies of Borowski *et al.* [31] and Fullone *et al.* [16], which find similar docked conformations, both of which also feature this salt bridge. The hydrophobic stabilisation provided by Phe121, Phe195 and Phe196, identified experimentally from the mutational analysis of Fullone *et al.*, is also observed. In the case of THR, the substrate’s position changes radically during molecular dynamics simulations, as the γ -OH group finds new hydrogen bonding partners in Glu102, the ferrous chloride and finally the oxoglutarate tail. The adoption of the stable hydrogen bond between the threonyl hydroxyl group and the oxoglutarate tail allows the threonyl ammonium-Glu102 salt bridge to be cleaved, which disrupts the hydrogen bonding network around the succinate tail. As a result, THR ends up in a very different position to ABA and NVA, which both largely maintain their docked conformations. The MD simulations with Fe(IV) active-site complexes again showed different positions for THR, depending on the hydrogen bonding partner of the γ -OH group, whilst ABA and NVA tended to keep their docked conformations.

The pentacoordinate species **D1** and **D2** generally had low values of d_{OH} and d_{CCl} . This is compatible with the findings of the spectroscopic study of Wong and co-workers [26], which suggest that it is one of these isomers that reacts. Interestingly, however, these results disagree with the literature in terms of the mechanism of isomer formation. Wong and co-workers propose that NVA forms the ammonium-Glu102 salt bridge on docking, but that during oxygen activation the salt bridge is broken and the ammonium group forms a hydrogen bond to the peroxy bridged structure, holding the oxoferryl oxygen *trans* to His235 and forming the **D2** isomer (Figure 4.14). In the case of THR, they propose that the γ -OH group retains its hydrogen bond to the carboxyl group of Glu102, which prevents the breaking of the salt bridge and allows the formation of the **D1** isomer. In the simulations reported in this thesis, however, this salt bridge is maintained in NVA, even in the presence of a potential nearby hydrogen bonding partner (the Fe=O oxygen in **D2** and **D3**). In THR, however, it is the very presence of the γ -OH group that *facilitates* the breaking of the salt bridge.

These forcefield calculations provide a better picture of the interactions of SyrB2 with its substrates, and should guide and inform further calculations of the mechanism.

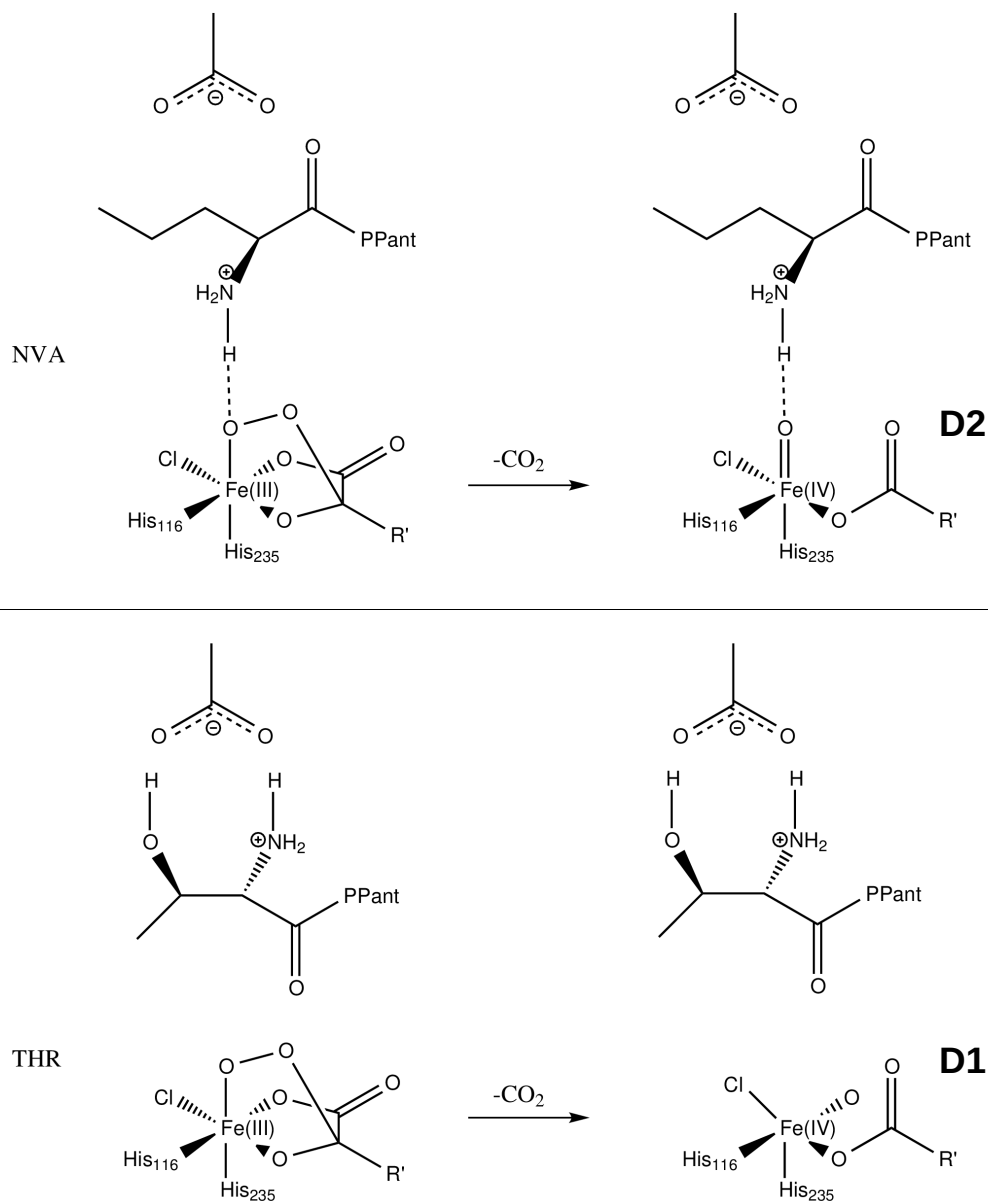


Figure 4.14: Mechanisms of Fe(IV) isomerisation proposed in literature [26]. Wong *et al.* suggest that interactions between NVA's ammonium group and a peroxy-bridged intermediate causes the formation of **D2**, whereas in THR's reaction the ammonium maintains its salt bridge to Glu102, allowing the formation of **D1**

Chapter 5

Mechanistic Calculations

5.1 Introduction

Armed with structures for the protein-substrate complexes and a better understanding of the limitations of the available methods, this chapter addresses mechanistic calculations. Proceeding from species **C**, the oxygen activation (**C–D**), hydrogen abstraction (**D–E**) and radical rebound (**E–F** and **E–F'**) steps (Figure 5.1) are modelled. These simulations are carried out both to identify which of the oxygen activation products **D1–D5** (Figure 1.7) is formed, and to explain the origin of the reaction's selectivity.

A number of prior studies, both experimental and computational, have attempted to address the mechanism of SyrB2, and studies of other NHFe enzymes also give insights into reactivity.

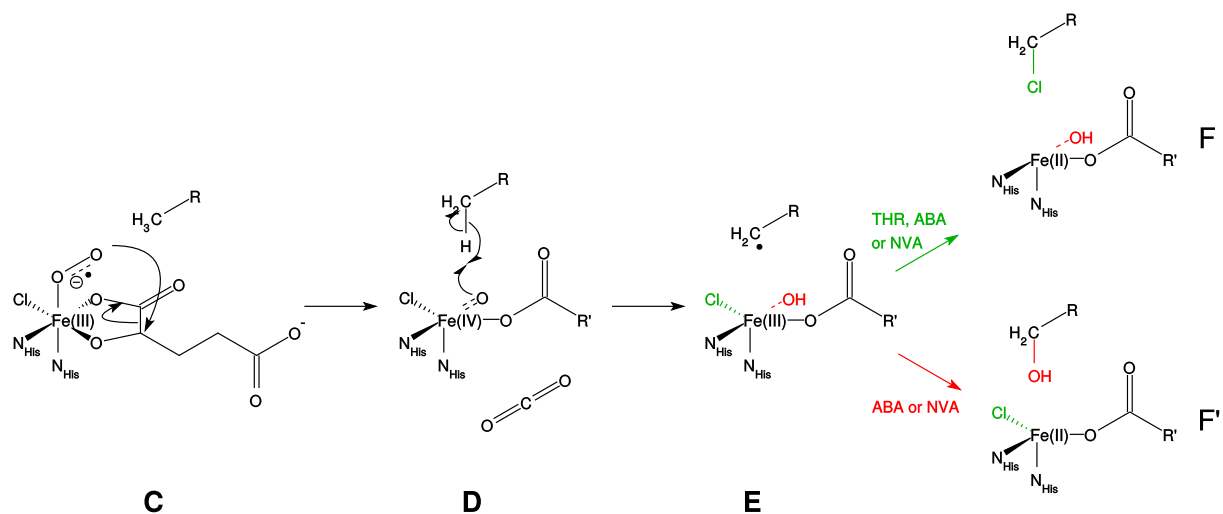


Figure 5.1: Oxygen activation, hydrogen abstraction, and radical rebound steps in the reaction of SyrB2 with ABA, NVA and THR.

5.1.1 Oxygen Activation

To date, only two studies [26, 30] have addressed oxygen activation in SyrB2. However, the subject of oxygen activation in NHFe enzymes has been studied extensively in other NHFe enzymes, particularly those with a “facial triad”-coordinated iron complex of the type shown in the upper panel of Figure 5.2 (FT-NHFe enzymes). From these studies, two broad, divergent “families” of mechanism have emerged [2, 110], one using the B3LYP functional, and one using the “spectroscopically calibrated” functional of Schenk and co-workers [77]. The functional-dependence of the mechanisms is due at least in part to the presence of multiple low-lying excited states for the oxygen adduct **AC** and the oxoferryl species **AD** (the notation of this thesis differentiates the intermediates of these reactions from the corresponding intermediates in SyrB2’s reaction with the use of the preceding letter **A**.).

Mechanisms from the B3LYP “family” have been proposed on multiple spin surfaces. On the quintet surface, the “B3LYP consensus” mechanism (Figure 5.2, Mechanisms 1 and 3) proceeds from $^5\mathbf{AC}$ to a peroxy-succinate complex $^5\mathbf{AH}$, either directly [37] (Mechanism 1) or through a peroxy-ketoacid intermediate $^5\mathbf{AG}$ with very little iron-oxygen interaction [111, 112] (Mechanism 3). The Fe(II) centre is then oxidised to Fe(IV), cleaving the oxygen-oxygen bond and giving the product $^5\mathbf{AD}$. An intermediate $^5\mathbf{AI}$ for the step $^5\mathbf{AH}$ – $^5\mathbf{AD}$, corresponding to a single-electron oxidation (giving Fe(III) and a half O–O bond), was detected by Ye and co-workers [37], but the barrier to the second oxidation was very low. On the septet surface, O–O bond cleavage and decarboxylation are concomitant, yielding $^7\mathbf{AD}$ [37]. Spin crossover then yields the quintet ground state. Ye and co-workers, attempting to replicate Mechanism 1 or 3 on the triplet surface, found the bicyclic intermediate $^3\mathbf{AG}$ to be prohibitively high in energy, higher than the transition state for its formation on the other two surfaces.

The other family of mechanisms, developed by Solomon and co-workers using the BP86 functional with 10% HF exchange [113], starts on the triplet surface. Oxygen binding leads directly to the peroxy-bridged species $^3\mathbf{AG}$ (Mechanism 4). As the reaction proceeds, this crosses over to the quintet surface through a bicyclic spin crossing point, before losing CO₂ to give the peroxy-bridged structure $^5\mathbf{AH}$. Cleavage of the O–O bond leads to the oxoferryl species $^5\mathbf{AD}$ as in Mechanisms 1 and 3.

Due to the role of spin-state energetics, the reaction mechanism is heavily dependent on both functional and ligand environment. For this reason, the oxygen-activation energetics of FT-NHFe enzymes will be likely to be somewhat different to those of SyrB2, as the latter has a weak-field chloride ligand where the former have a medium-field carboxylate ligand. To date, there have only been two studies of oxygen activation in SyrB2. Wong

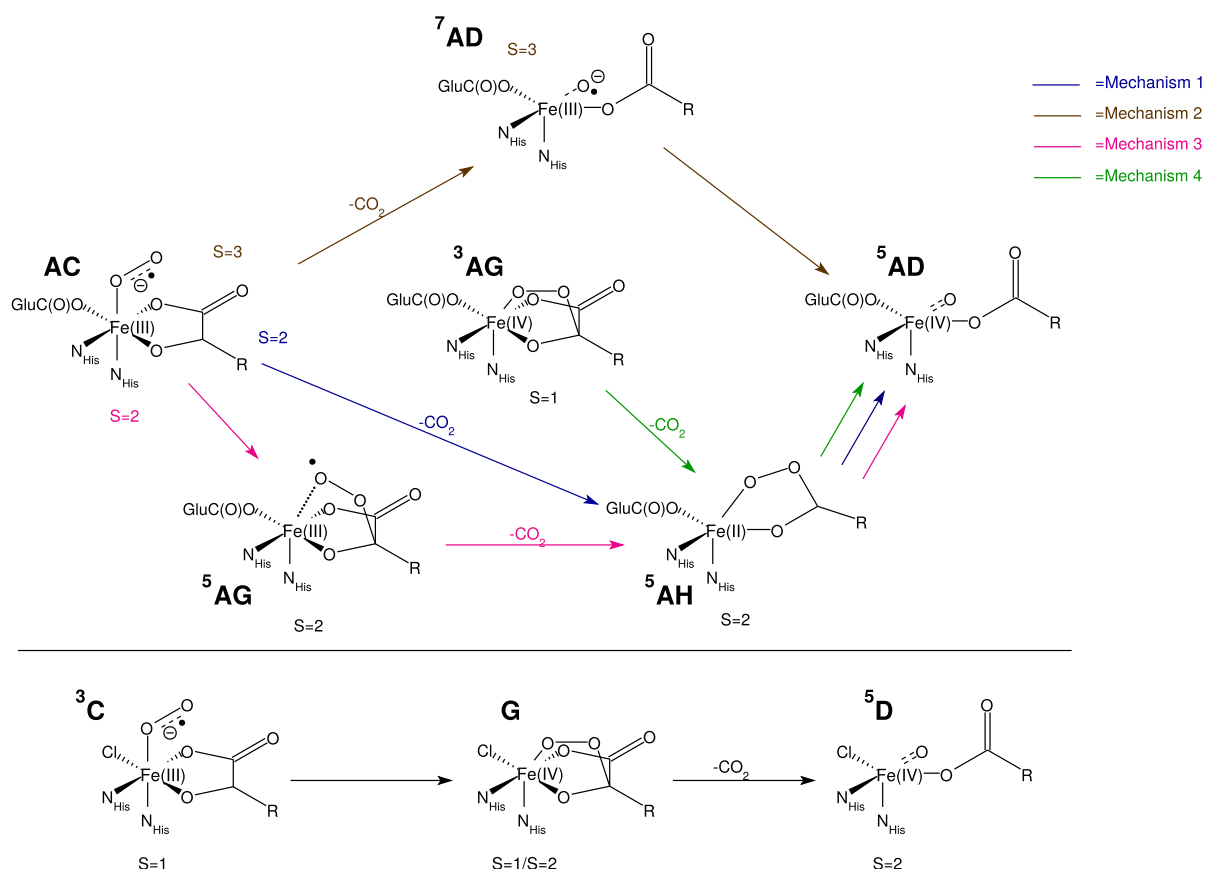


Figure 5.2: Proposed mechanisms of oxygen activation in NHFe enzymes. *Top*: FT-NHFe enzymes. **³AG** differs structurally from **⁵AG** in that the iron-dioxygen bond is considerably longer and weaker in **⁵AG**. *Bottom*: SyrB2 [26]

and co-workers [26], using BP86, follow a mechanism similar to Mechanism 4 (shown in Figure 5.2, bottom panel). In this mechanism, oxygen binding leads directly to the formation of an adduct of type **³C**. Next, the species **³G** was formed, which has a near-degenerate quintet state. Spin crossover to this quintet state led to O–O bond cleavage and concomitant decarboxylation, and ultimately to the oxoferryl species **⁵D1**. Kulik and co-workers [30], using the PBE functional with a Hubbard U correction, followed a similar mechanism. Initial oxygen binding led to the formation of an oxygen adduct of type **⁵C**, with the dioxygen O–O axis rotated to point the attacking oxygen away from the 2OG ketone carbon. When this oxygen is brought close enough to this carbon, an essentially barrierless transition to a bridged, bicyclic structure occurs. Exothermic decarboxylation and concomitant O–O bond cleavage lead to the product **⁵D2**.

5.1.2 Hydrogen Abstraction

As with oxygen activation, the mechanism of hydrogen abstraction in N^HFe enzymes has been studied extensively [36, 72, 114]. It is, however, somewhat better understood, not least because the oxoferryl intermediate from which hydrogen abstraction commences is experimentally isolable. The barrier to hydrogen abstraction has been shown to be dependent upon both the multiplicity and the direction of the approach of the incoming hydrogen relative to the axis of the oxoferryl bond. If the Fe–O–H angle is at or around 180°, the substrate donates an electron into a σ^* orbital of the oxoferryl bond (σ -pathway reactivity), whilst with values of this angle closer to 120° it is a π^* orbital that is attacked (π -pathway reactivity). On the quintet surface, the σ -pathway is favoured, whilst on the triplet surface the preference is reversed. The reason for the preference for σ -channel reactivity on the quintet surface is that, through antiferromagnetic coupling, this allows a greater number of unpaired electrons in the product than in the reactant, a phenomenon known as exchange enhancement [115]. By promoting an electron to the d_{z^2} orbital, reactivity through the σ -channel leads to six unpaired electrons in the radical product, whereas reactivity through the π -channel leads to only four (Figure 5.3). This “exchange enhancement” can have profound consequences for reactivity, with some systems [114] undergoing spin-crossover to take advantage of it. In SyrB2, however, it is believed that only the quintet surface is relevant to hydrogen abstraction [31].

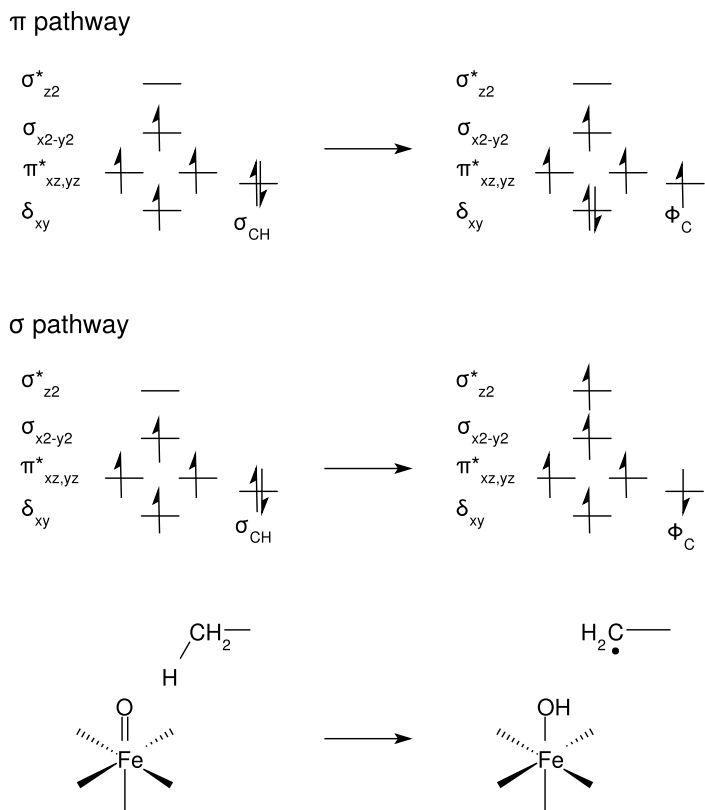


Figure 5.3: Exchange enhancement on the quintet surface. π -channel reactivity (top panel) allows only four unpaired electrons in the product, whereas σ -channel reactivity (central panel) allows a total of six.

5.2 Computational Details

QM/MM calculations were carried out using the ChemShell 3.6 environment [116, 117] to link Turbomole 6.4 [78, 79, 80] to ChemShell’s implementation of DL.POLY 2 [118]. A link-atom scheme was used, with electrostatic interactions between the QM and MM regions calculated using electrostatic embedding. Microiterative optimisation was used with the DL-FIND algorithm [119], using HDLC coordinates [63]. Transition states were located with the p-rfo algorithm [62].

QM calculations used the B3LYP functional with a dispersion correction (DFT-D3 [50]). All geometries were initially calculated at the def2-SVP level [81, 82], with stationary points reoptimised at the def2-TZVP level. All data in this chapter, unless otherwise stated, comes from these subsequent def2-TZVP-level calculations. MM calculations used the Amber ff03 forcefield [103] for the protein, with the TIP3P water model [101]. Charges for the substrate, iron complex and cofactors were the same as those used for MD with Fe(II) active site, as described in Section 4.2.2. However, most of these atoms fell within

the QM region, as a result of which their charges were made redundant by the use of electrostatic embedding.

Snapshots were taken from the equilibrated trajectories of the Fe(II) simulations. Snapshot choice is discussed in more detail in Section 5.3.1. During optimisation, the system was divided into four distinct regions: the outer region, the active region, the inner region, and the QM region. The outer region is unable to move, and is present to provide environment effects to the active region. In each case it was chosen as the entire protein, as well as any water molecules within a 23 Å radius of the central iron (excluding any atoms in the active region). The active region is able to move, and was chosen as all residues within an 8 Å shell of the inner region residues. The inner region, which is the region for which optimisation steps are kept to a minimum during microiterative optimisation (see Section 2.5), was chosen as all residues that included QM atoms. The QM region was chosen to include the iron complex and the substrate, as well as the Glu102–substrate amino salt bridge, to allow charge transfer. It consisted of Fe and all first-shell ligands (with histidines truncated to imidazoles and 2-oxoglutarate truncated to 2-oxopropanoate), as well as the substrate head (the amino acid section of the substrate-pantetheine complex) and the carboxylate group of Glu102. Since, during MD simulations of THR, the Glu102–substrate amino salt bridge was water-mediated, the two bridging water molecules were also included in the QM region for calculations involving this substrate. The inner and QM regions for THR are shown in Figure 5.4. For this system, the sizes of the QM region, inner region, active region and outer region are 62, 106, 1550 and 6862 atoms respectively.

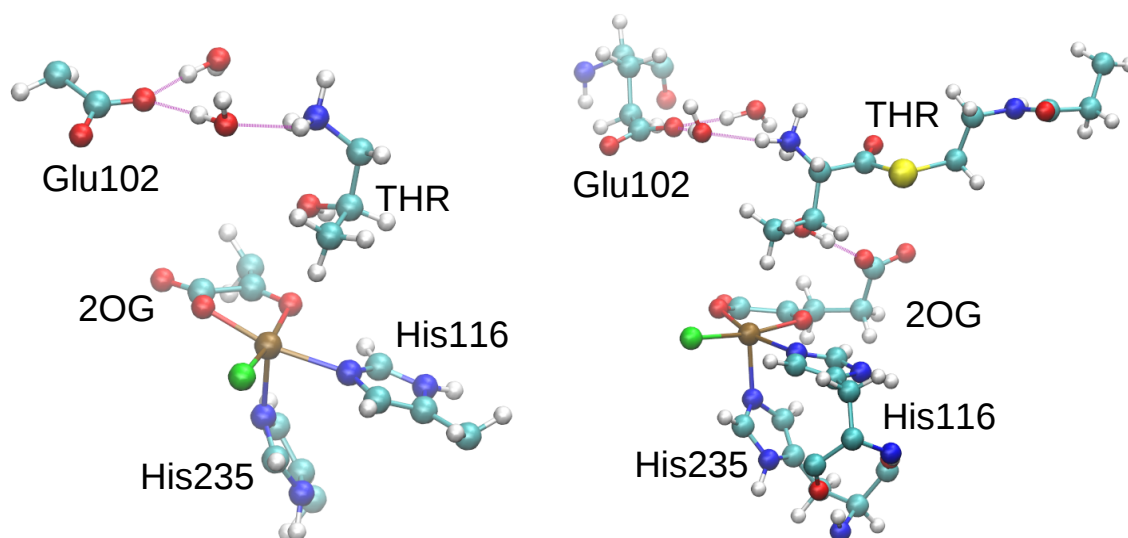


Figure 5.4: Left QM region and right: inner region of THR snapshot. Note the water mediated Glu102-threonyl ammonium salt bridge, included to allow proton transfer.

Reactions were modelled by incrementing the difference of distances of reacting atoms. This was carried out by optimisation using difference-of-distance restraints of 2.8×10^6 kJ mol⁻¹ nm⁻² at the required increments.

5.3 Results and Discussion

5.3.1 Oxygen Activation

Structure of the Oxygen Adduct

QM/MM calculations for each substrate commenced from a representative snapshot from the MD trajectory of the protein with Fe(II) active site and that substrate (corresponding to species **B**). All structures chosen were within the standard deviation of the backbone RMSD for that simulation. The structure selected for THR had a hydrogen bond between the γ -OH group and the succinate tail (see Section 4.3.2). The snapshot chosen for NVA had a $C^\alpha - C^\beta - C^\gamma - C^\delta$ torsion of 82° , which kept the substrate head sufficiently far from the Fe complex to allow oxygen binding without major conformational change. The structures were optimised at quintet multiplicity. The Fe complexes and immediate environments (including the substrate heads) did not undergo any significant conformational change from the MD snapshot. Geometric parameters are available in Table 5.7.

Next, dioxygen was placed between Fe and the 2OG ketone carbon C_K (Figure 5.5) and the new structures optimised, at all three of the triplet, quintet and septet spin states for each substrate (leading to species **C**). In the presence of THR and ABA, adducts were stable at all three multiplicities, although for THR the triplet adduct was only stable at the def2-SVP basis level. In the presence of NVA, only the quintet species was stable, and had a different structure, both electronically and geometrically, to the quintet adducts for the other two substrates. Attempts to find triplet and septet adducts in the presence of NVA failed, with the oxygen moving away from the iron complex.

These complexes are directly comparable to the nitrosyl adducts isolated by Martinie and co-workers [28], allowing comparison of the Fe-H distances to those reported from their hyperfine sublevel correlation spectroscopy experiments. The distances between Fe and the reactive hydrogens for ABA and THR match those reported by Martinie and co-workers very closely (see Table 5.1). Those for NVA do not, possibly due to the conformation of the substrate head. During MD simulations, multiple conformations of the substrate head were observed, and at a later stage in the reaction, a rotation around the $N - C^\alpha - C^\beta - C^\gamma$ torsion occurs, which brings the norvaline head closer to the reactive oxygen (see Section 5.3.2).

	NVA C^γ	NVA C^δ	THR C^γ	ABA C^γ
Fe-H (Å) [28]	3.7	3.4	4.2	3.7
Fe-H (Å) (This thesis, S=1)	—	—	—	3.9
Fe-H (Å) (This thesis, S=2)	3.2	5.1	4.1	3.9
Fe-H (Å) (This thesis, S=3)	—	—	4.2	3.9

Table 5.1: Comparison of model complexes to experimentally validated interatomic distances

For the substrates for which it was stable, the septet was the ground state (Table 5.2). The subject of the ground state, and indeed the electronic structure, of \mathbf{C} is one upon which the literature is somewhat divided. As discussed in Section 5.1.1, studies of oxygen activation have used different ligand environments and different functionals, which will inevitably affect the spin energetics. The two previous studies that model this species in SyrB2 [30, 26] find different ground states to one another, even when both use a pure functional. Studies of FT-NHFe enzymes are similarly discordant. Some examples of these are shown in Table 5.2.

Analysis of the spin populations of the various adducts revealed a number of different electronic configurations. The septet and quintet states of THR’s complex \mathbf{C} , and all three states of ABA’s, had a simple superoxide structure, with a single unpaired electron delocalised across the dioxygen π -bond, and the remaining 1, 3 or 5 unpaired electrons (for the triplet, quintet or septet states respectively) ferromagnetically coupled to it from the iron d-orbitals (see Figure 5.5). However, in the case of NVA’s ${}^5\mathbf{C}$, a similar superoxo structure was observed, but with a total of *six* unpaired electrons—five in the Fe d orbitals coupled antiferromagnetically to one in the dioxygen π -bond. THR’s ${}^3\mathbf{C}$ had the most divergent electronic structure, with an Fe(II) centre and a neutral oxygen diradical, again with a total of six unpaired electrons. Interestingly, this case was lower in energy than the quintet, whilst for ABA the triplet had the highest energy of any spin state.

These differences in electronic structure are borne out in the geometries (Table 5.3 and Figure 5.6). The two comparable quintet structures (those of ABA and THR), had very similar geometries. The same is true of the septet species for these two substrates. For these two sets of structures, the transition from quintet to septet is marked by an elongation of the two Fe – N_{His} bonds from approximately 2.0 Å to approximately 2.1 Å. The Fe – O_P bond is also considerably shorter in these two quintet structures (at around 1.95 Å) than in the corresponding septet structures, in which it has values in excess of 2.1 Å. The structure of the quintet adduct of NVA more closely resembled that of the adducts in the septet states for the other two species, with Fe – N_{His} bond lengths at

Model Class	Parameter	Method	S=1	S=2	S=3
SyrB2 + THR* (this thesis)	ΔE (kJ/mol)	B3LYP	44	50	0
	$\rho_u(\text{Fe})$		3.37	2.91	4.17
	$\rho_u(\text{O}_\text{P})$		-0.73	0.38	0.63
	$\rho_u(\text{O}_\text{D})$		-0.81	0.52	0.54
SyrB2 + ABA (this thesis)	ΔE (kJ/mol)	B3LYP	50	41	0
	$\rho_u(\text{Fe})$		1.09	2.93	4.14
	$\rho_u(\text{O}_\text{P})$		0.49	0.37	0.66
	$\rho_u(\text{O}_\text{D})$		0.50	0.47	0.63
SyrB2 + NVA (this thesis)	ΔE (kJ/mol)	B3LYP	—	n/a	—
	$\rho_u(\text{Fe})$		—	4.15	—
	$\rho_u(\text{O}_\text{P})$		—	-0.22	—
	$\rho_u(\text{O}_\text{D})$		—	-0.49	—
SyrB2 [26]	ΔE (kJ/mol)	BP86	0	23	40
SyrB2 [30]	ΔE (kJ/mol)	PBE+U	13	0	—
SyrB2 [30]	ΔE (kJ/mol)	PBE	4	0	—
FT-NHFe [37]	ΔE (kJ/mol)	B3LYP	0	5	2
FT-NHFe [112]	ΔE (kJ/mol)	B3LYP	32	45	0

Table 5.2: Spin populations and relative spin-state energies of the oxygen adduct, from this study and from literature. ρ_u refers to the spin density $\rho_\alpha - \rho_\beta$ from the Mulliken spin population of the atom in question. O_P and O_D refer to the atoms of the bound dioxygen that are proximal and distal to the iron, respectively. *Data for THR/S=1 comes from def2-TZVP calculation using coordinates from def2-SVP optimisation

around 2.1 Å and an Fe – O_P bond length of 2.04 Å, which is in accord with its electronic structure, which also resembles the septet for the other two substrates. The triplet adduct of THR, which is electronically unlike any of the other species, has an equally aberrant structure.

Oxygen Activation

Decarboxylation was modelled on all available spin surfaces in the presence of each substrate. The reaction was driven by gradually increasing the difference of O–O and C–O distances by 0.05 Å increments.

On the quintet surface, the scans, whose energy profiles are shown in Figure 5.7 and Figure 5.8, tended to follow one of the “B3LYP consensus” mechanisms (Figure 5.2, Mechanism 1 or 3). THR and ABA both progressed from the oxygen adduct to a bicyclic inter-

Parameter	THR			ABA			NVA
	S=1*	S=2	S=3	S=1	S=2	S=3	S=2
Fe–N _{His235} (Å)	2.02	1.99	2.05	1.98	2.00	2.09	2.12
Fe–N _{His116} (Å)	2.11	2.02	2.12	2.00	2.01	2.11	2.13
Fe–Cl (Å)	2.32	2.31	2.27	2.24	2.29	2.24	2.28
Fe–O _P (Å)	2.01	1.95	2.22	1.95	1.94	2.13	2.04
Fe–O _{A1} (Å)	2.06	1.98	2.06	1.95	1.99	2.05	2.02
Fe–O _K (Å)	2.30	2.29	2.08	1.99	2.32	2.28	2.25
O–O (Å)	1.24	1.28	1.28	1.29	1.29	1.28	1.28
N _{His235} –Fe–O _P (°)	169	164	157	178	171	164	158

Table 5.3: Geometric parameters of the oxygen adduct **C**. *THR/S=1 bond lengths come from optimisation at def2-SVP basis level

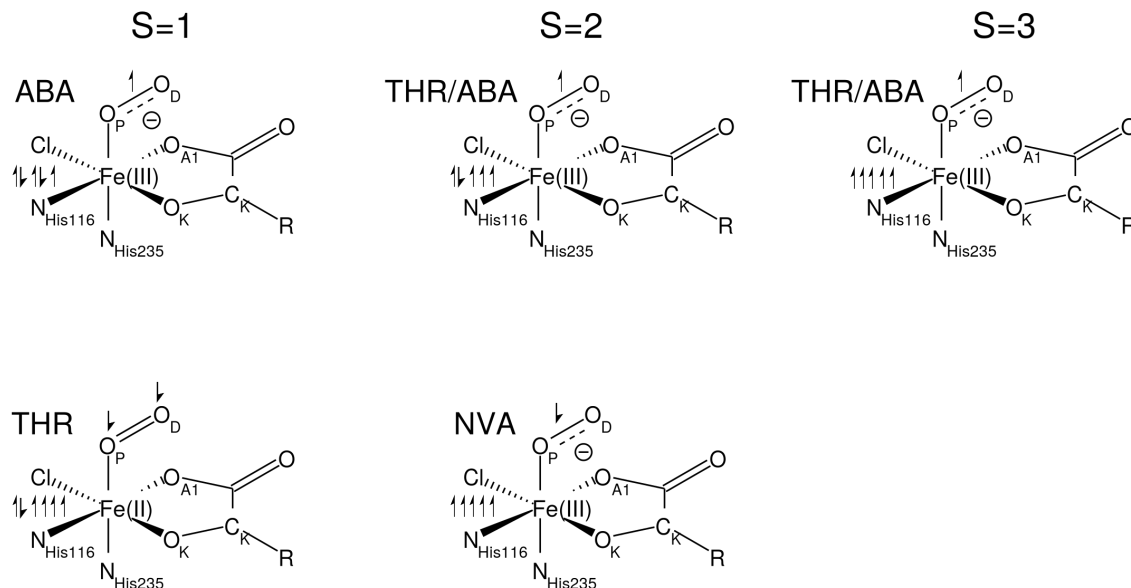


Figure 5.5: Electronic structures of the oxygen adduct **C**

mediate **G**, similar to **AG** (see Figure 5.12, left panel), with peroxy-succinate character (the Fe – O_P bond was lengthened considerably, see Table 5.4). In NVA this intermediate was not detected. Following a highly exothermic decarboxylation step, all three substrates formed a peroxy-bridged intermediate **H** (Figure 5.12, central panel). The O–O bond is then cleaved one electron at a time, yielding first the half-bond intermediate **I** (Figure 5.12, right panel), resembling **AI**, then the Fe(IV)-oxo species **D**. In each case this was

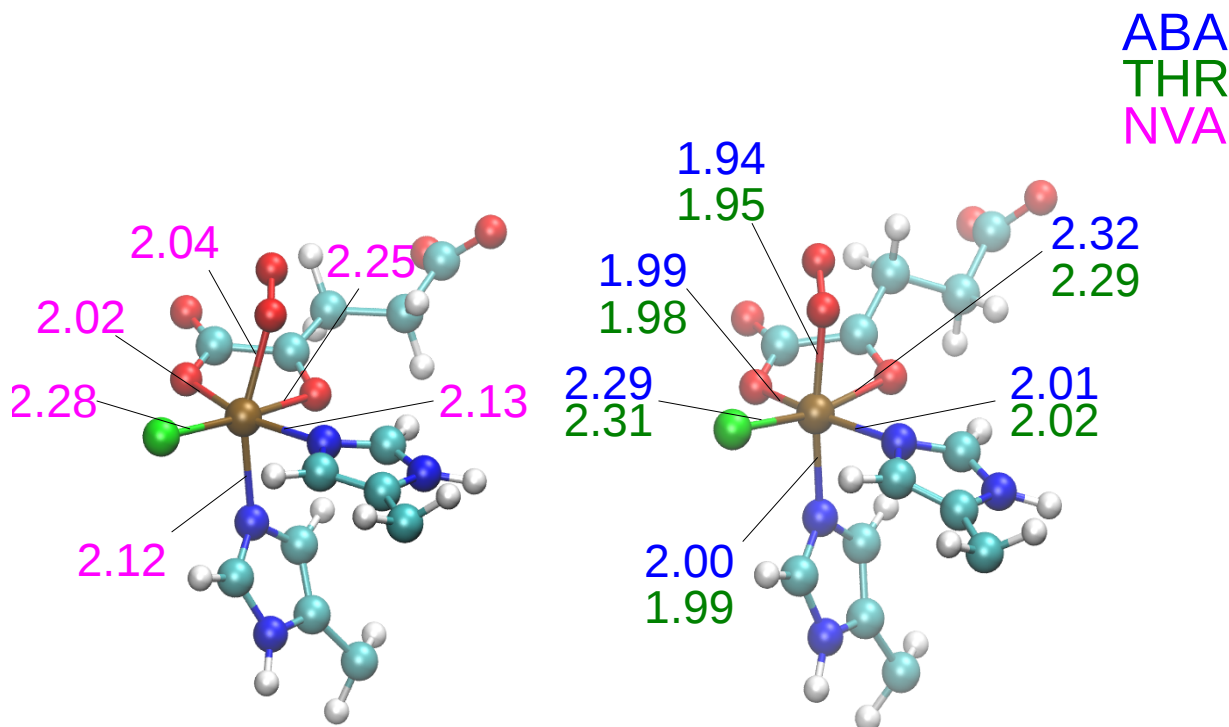


Figure 5.6: Oxygen adduct (**C**) for NVA (left) and THR and ABA (right)

the same isomer, **D1** (Figure 5.14).

Scans on the other two spin surfaces are shown in Figures 5.10 and 5.11 for ABA and THR respectively. Energy profiles for the latter are shown in Figure 5.15. The two scans on the septet surface underwent a similar mechanism to those on the quintet surface, with the notable exception that the bicyclic intermediate **G** was not observed, and a short-lived intermediate **J** (Figure 5.9) was formed immediately after decarboxylation. **J** retains a charge on the CO₂, corresponding to a CO₂ anion coordinated to an Fe(III) centre. Following decarboxylation, species **H** and **D** were observed, with similar structures to their analogues on the quintet surface, the end product again being **D1**.

The two scans on the triplet surface were very different from one another, as might be expected from the radically different electronic structure of ³**C** formed in the presence of THR and ABA. The THR triplet scan underwent a mechanism very similar to that on the quintet surface, with **G** and **H** both observed prior to the formation of **D1**. For the ABA triplet scan, however, the mechanism was entirely different. No intermediate was detected between **C** and **D**, with the scan rising to over 151 kJ/mol before forming **D2** (Figure 5.13). This was the only scan to lead to a structure other than **D1**.

The structure of **D1** from the three scans on the quintet surface is shown in Figure 5.14. Hydrogen bonds from the terminal hydrogens of Arg254 hold the ferryl and carboxyl

Oxygen Activation on the Quintet Surface

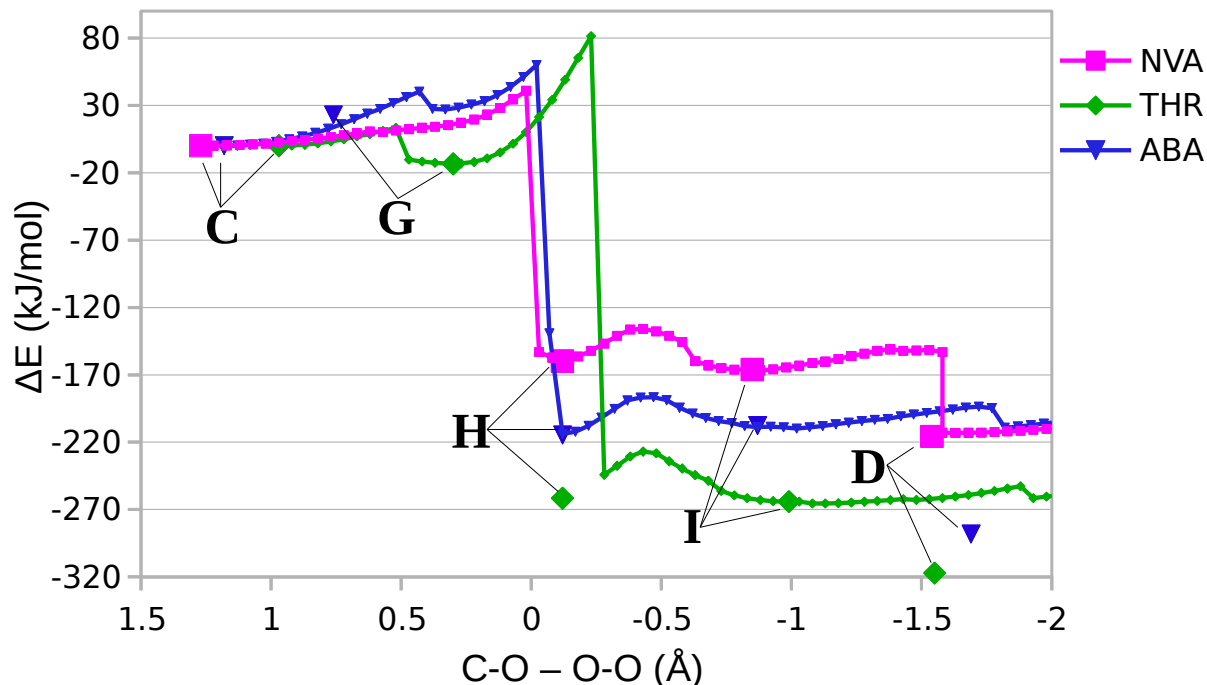


Figure 5.7: Energy profile of oxygen activation on the quintet surface. Stationary points (optimised without restraints) are shown with larger symbols. All points plotted are at the def2-SVP basis level

oxygens that arise from oxygen activation. This structure was the same as that obtained from the end point of the two septet scans, as well as the triplet scan in the presence of THR. NVA's $C^\alpha - C^\beta - C^\gamma - C^\delta$ torsion had a value of 77° , which was a minor state detected in the NVA/**D1** MD simulations. d_{OH} and d_{CCl} were generally significantly shorter than those obtained in the MD simulations of Section 4.3.3 (see Table 5.5), generally to the extent of being outside the standard deviation. It is possible that this is because in the structure obtained from oxygen activation the substrate has not yet fully equilibrated to the newly formed oxoferryl environment, but that during MD simulations of **D1** the hydrophobic substrate head was able to separate from the hydrophilic oxoferryl unit.

Spin-state separations of the oxoferryl species are shown in Table 5.6. In spite of the similar structure of **D** at different multiplicities for each substrate, these values show some differences to those calculated for Models **1a**, **1b** or **3b** (see Chapter 3). For THR, $^7\mathbf{D}$ has an unexpectedly low energy relative to the other multiplicities. This is due to a minor change in the hydrogen bonding environment some distance from the oxoferryl complex, caused by a water molecule adopting a new conformation with a different hydrogen-bonding partner. The different isomerism of ABA's $^3\mathbf{D}$ species (**D2** rather than

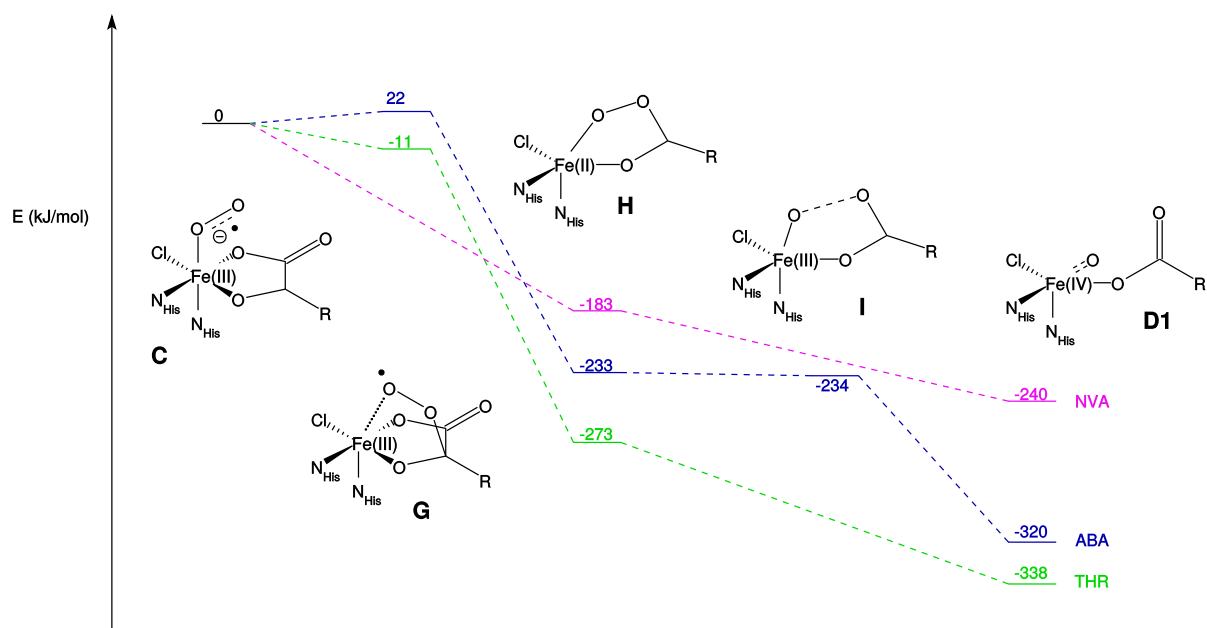


Figure 5.8: Energy profile of oxygen activation for the three substrates on the quintet surface at the def2-TZVP level. Attempts to reoptimise **I** for THR and NVA at this basis level led to the cleavage of the $\text{O}_\text{P}\text{--}\text{O}_\text{D}$ bond and the formation of **D**

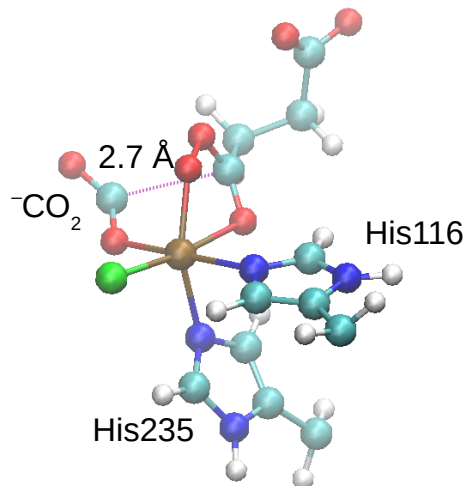


Figure 5.9: Species **J**, observed during oxygen activation on the septet surface. At 2.7 Å, the C–C bond has been severed, but the CO_2^- unit retains its charge

D1) significantly raises the energy of this species relative to ^5D and ^7D .

The spin populations of the ^5D and ^7D are similar to those of Model **1a** (Figure 3.3). The three structures of ^5D each have a spin density at or around 3.1 on the central iron and around 0.6 on the oxoferryl oxygen, the spin populations of these atoms for the quintet

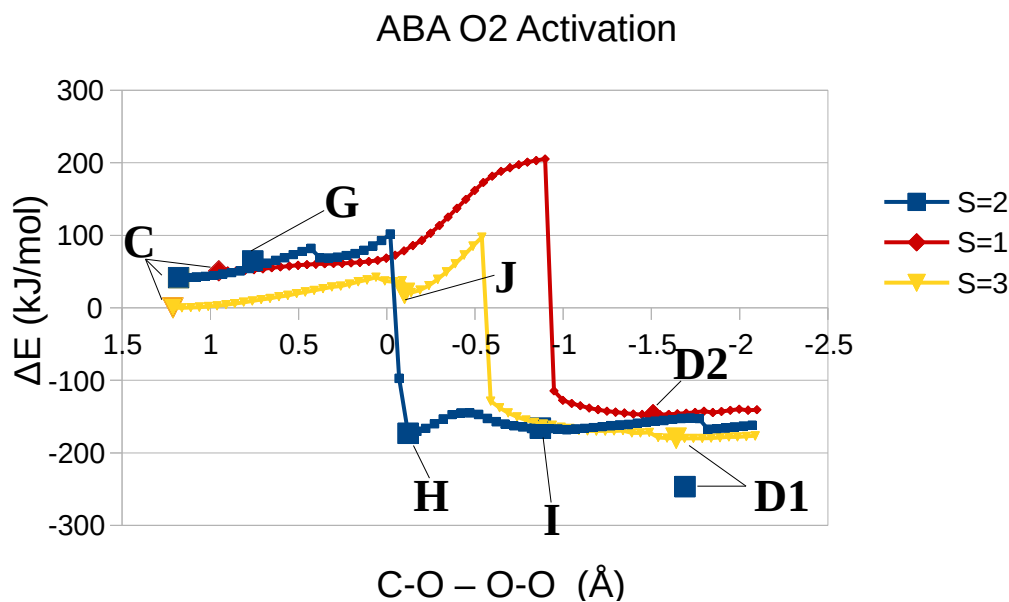


Figure 5.10: Comparison of energy profiles of oxygen activation on the triplet, quintet and septet surfaces for the substrate ABA. Stationary points (optimised without restraints) are shown with larger symbols. All points plotted are at the def2-SVP basis level

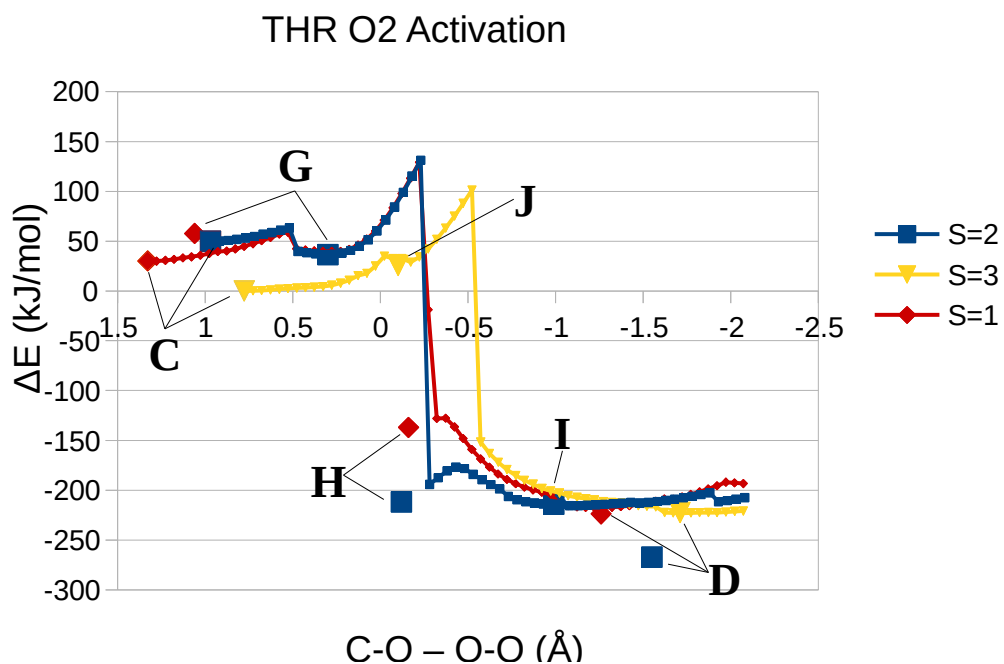


Figure 5.11: Comparison of energy profiles of oxygen activation on the triplet, quintet and septet surfaces for the substrate THR. Stationary points (optimised without restraints) are shown with larger symbols. All points plotted are at the def2-SVP basis level

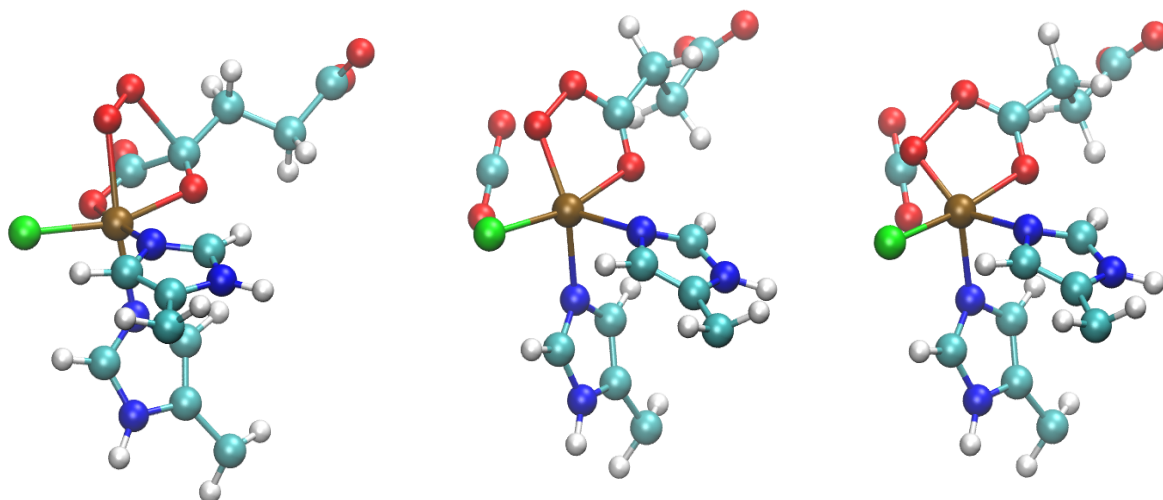


Figure 5.12: Oxygen activation intermediates (from ABA scan) Left: **G** Centre: **H** Right: **I**

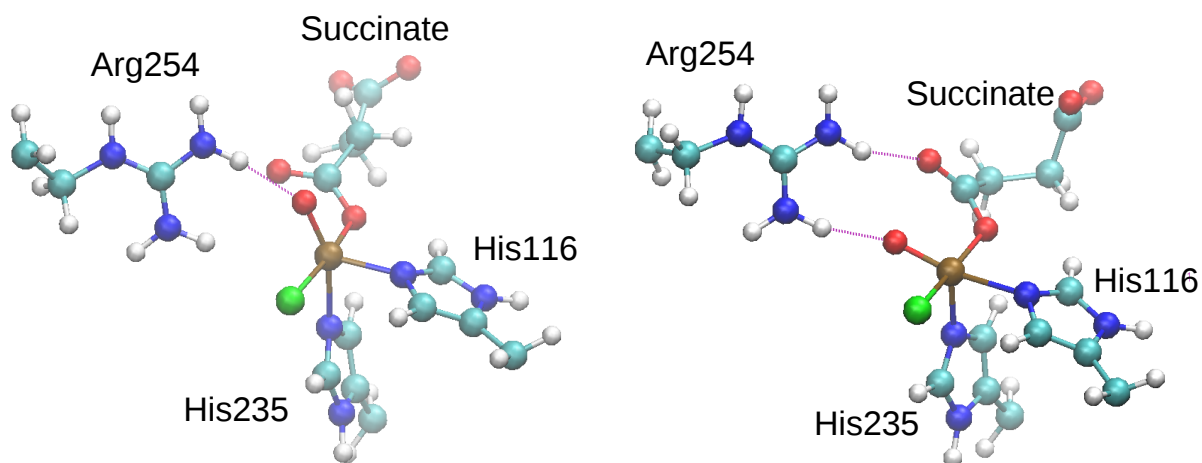


Figure 5.13: Products of oxygen activation on the triplet surface. Left: $^3\text{D2}$, formed for ABA, Right $^3\text{D1}$, formed for THR

state of Model **1a**. Similarly, the two septet species each had a spin density of around 4.1 on the central iron, with around 1.1 on the oxoferryl oxygen, compared with 4.1 and 1.4 for the iron and oxoferryl oxygen, respectively, in the septet state of Model **1a**. The two triplet structures, however, had rather different structures to one another. The structure formed in the presence of ABA more closely resembled the spin populations of Model **1a** (with spin densities of 1.6 and 0.46 on the Fe and oxoferryl oxygen, respectively, compared to 1.2 and 0.8 for Model **1a**). That formed in the presence of THR, however, had a spin population of 2.83 on iron, with -0.83 on the oxoferryl oxygen, which is a starkly different

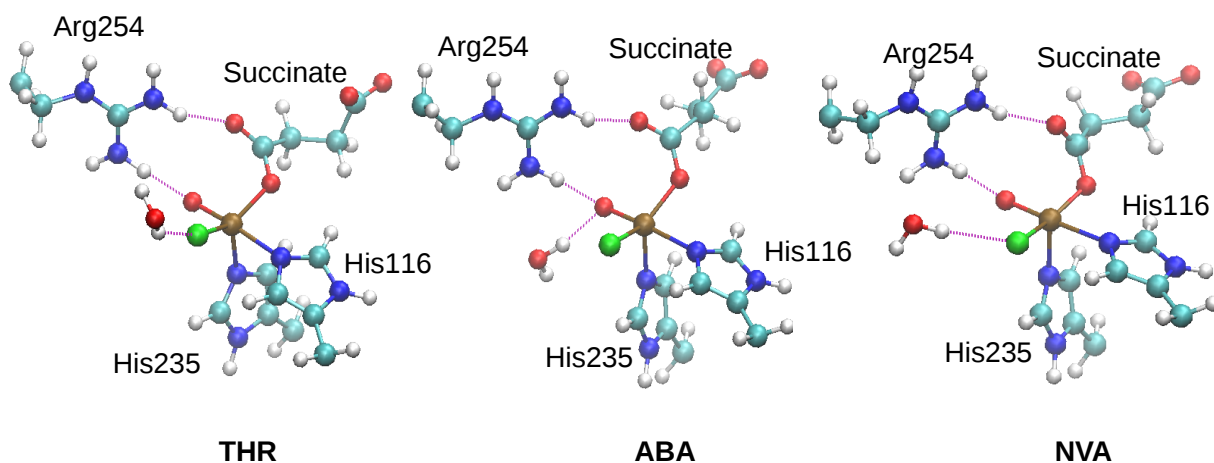


Figure 5.14: Oxoferryl species $^5\mathbf{D1}$ formed from oxygen activation on the quintet surface.

electronic configuration to that of triplet Model **1a**.

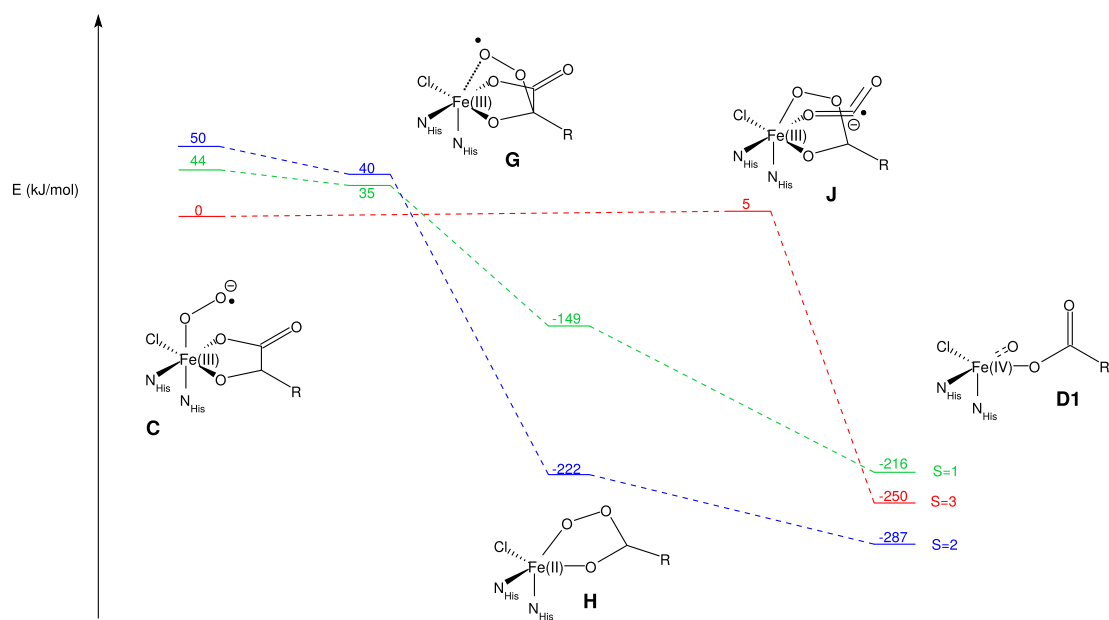


Figure 5.15: Oxygen activation in the presence of THR.

In summary, whilst these calculations uncovered numerous electronic structures of **C**, oxygen activation scans from all but one of these led to the same isomer of **D**, **D1**. The formation of **D1** is in accord with the findings of Wong and co-workers [26], both spectroscopic and computational, although Wong and co-workers propose that **D2** may

be formed in the presence of NVA. On the basis of this consensus, hydrogen abstraction scans were carried out on the structures of $^5\mathbf{D1}$.

	Parameter	⁵ C	⁵ G	⁵ H	⁵ D	³ C	³ G	³ H	³ D	⁷ C	⁷ D
THR	Fe-N _{His235} (Å)	1.99	2.09	2.15	2.10	2.05*	1.98	2.27	1.94	2.05	2.07
	Fe-N _{His116} (Å)	2.02	2.01	2.13	2.10	2.12*	1.98	2.02	2.10	2.12	2.16
	Fe-Cl (Å)	2.31	2.25	2.32	2.25	2.27*	2.27	2.28	2.29	2.27	2.32
	Fe-O _P (Å)	1.95	2.40	2.05	1.63	2.22*	1.82	1.91	1.60	2.22	1.90
	Fe-O _K (Å)	2.29	1.91	2.26	1.93	2.08*	1.84	2.01	1.96	2.08	1.92
	O _P -O _D (Å)	1.28	1.30	1.44	n/a	1.28*	1.41	1.48	n/a	1.28	n/a
	O _D -C _K (Å)	2.25	1.60	1.32	1.22	2.57*	1.45	1.91	1.22	1.93	1.22
	N _{His235} -Fe-O _P (°)	164	157	114	95	157*	171	165	111	157	106
ABA	Fe-N _{His235} (Å)	2.00	2.11	2.12	2.07	1.98	—	—	2.08	2.09	2.07
	Fe-N _{His116} (Å)	2.01	1.99	2.11	2.10	2.00	—	—	1.96	2.11	2.12
	Fe-Cl (Å)	2.29	2.26	2.35	2.26	2.24	—	—	2.33	2.24	2.36
	Fe-O _P (Å)	1.94	2.27	2.02	1.64	1.95	—	—	1.59	2.13	1.87
	Fe-O _K (Å)	2.32	2.01	2.33	1.94	1.99	—	—	1.99	2.28	1.95
	O _P -O _D (Å)	1.29	1.28	1.44	n/a	1.29	—	—	n/a	1.28	n/a
	O _D -C _K (Å)	2.47	2.07	1.32	1.22	2.23	—	—	1.23	2.43	1.23
	N _{His235} -Fe-O _P (°)	171	169	139	93	178	—	—	156	164	101
NVA	Fe-N _{His235} (Å)	2.12	—	2.10	2.05	—	—	—	—	—	—
	Fe-N _{His116} (Å)	2.13	—	2.11	2.12	—	—	—	—	—	—
	Fe-Cl (Å)	2.28	—	2.37	2.29	—	—	—	—	—	—
	Fe-O _P (Å)	2.04	—	2.00	1.63	—	—	—	—	—	—
	Fe-O _K (Å)	2.25	—	2.33	1.95	—	—	—	—	—	—
	O _P -O _D (Å)	1.28	—	1.45	n/a	—	—	—	—	—	—
	O _D -C _K (Å)	2.47	—	1.32	1.22	—	—	—	—	—	—
	N _{His235} -Fe-O _P (°)	158	—	132	96	—	—	—	—	—	—

Table 5.4: Geometric parameters of the species formed during oxygen activation. Parameters for **I** and the **J** are not shown, but are available in the appendix in Table A.1.

*THR/³C coordinates calculated at def2-SVP level

Case	d_{OH} (QM/MM)	d_{OH} (MD \bar{x})	d_{OH} (MD σ)	d_{CCl} (QM/MM)	d_{CCl} (MD \bar{x})	d_{CCl} (MD σ)
NVA (C^γ)	3.65	4.51	0.45	3.64	4.04	0.16
NVA (C^δ)	5.58	3.74	1.06	3.74	3.98	0.27
THR (Conf. 1)	3.01	4.41	0.45	3.64	4.50	0.53
THR (Conf. 2)	2.80	4.17	0.36	3.65	4.36	0.43
ABA	4.17	4.59	0.52	3.74	4.22	0.27

Table 5.5: Comparison of d_{OH} and d_{CCl} between MD simulations of **D1** and structures of **⁵D1** obtained from oxygen activation on the quintet surface. \bar{x} refers to the mean and σ to standard deviation

Parameter	THR			ABA			NVA		
	$S = 1$	$S = 2$	$S = 3$	$S = 1$	$S = 2$	$S = 3$	$S = 1$	$S = 2$	$S = 3$
ΔE (kJ/mol)	55	0	37	107*	0	69	—	n/a	—
$\rho_u(\text{Fe})$	2.83	3.1	4.12	1.61*	3.16	4.12	—	3.14	—
$\rho_u(\text{O}_\text{P})$	−0.83	0.64	1.11	0.46*	0.55	1.11	—	0.63	—

Table 5.6: Spin-state separations and spin populations of the oxoferryl complexes **D**. *The ABA/ $S=1$ scan ended with the isomer **D2**, whilst all other scans ended with **D1**

5.3.2 Hydrogen Abstraction

Next, hydrogen abstraction from the structures of $^5\mathbf{D1}$ was modelled, by decreasing the difference of the O–H and C–H distances for the reacting atoms. In addition to the three structures obtained from oxygen activation (taken from the scans on the quintet surface), a second structure was prepared for THR to reflect the second conformation observed during MD simulations (see Section 4.3.3). In the conformation obtained from the oxygen activation scans, the γ -hydroxyl group has a hydrogen bond to the succinate tail (Conformation 1), whilst during MD simulations a second stable conformation (Conformation 2) was observed, in which the γ -hydroxyl group is instead hydrogen-bonded to the succinate head (Figure 5.16). A structure for \mathbf{D} in Conformation 2 was generated by rotating the γ -OH bond of Conformation 1 and reoptimising. This new structure was 18 kJ/mol higher in energy. A total of 5 hydrogen abstraction scans were run—four from the terminal carbon of each of the four structures described above, and one from C^γ of NVA. As species \mathbf{D} is known from Mössbauer spectroscopy to have a quintet ground state [27], and spin crossover is not considered likely to occur during hydrogen abstraction and radical rebound [31], all subsequent reaction scans are calculated on the quintet surface.

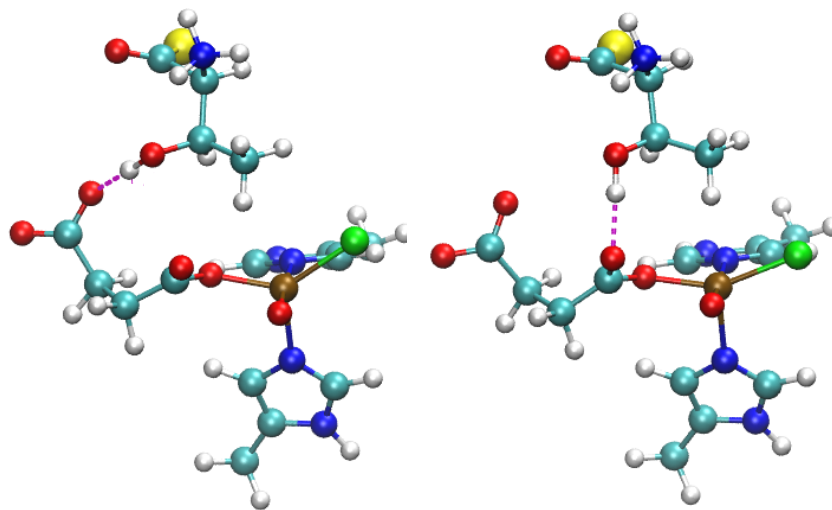


Figure 5.16: THR in Conformation 1 (left) and 2 (right)

Scans for abstraction of a hydrogen from the terminal carbon of ABA, NVA and THR in Conformation 1 are shown in Figure 5.20, whilst Figure 5.26 shows an energy-level diagram for these reactions. Abstraction from NVA’s C^δ yields an essentially degenerate secondary minimum before the transition state (at a difference of distances of 1.7 Å),

which corresponds to a rotation around the $\text{N} - \text{C}^\alpha - \text{C}^\beta - \text{C}^\gamma$ torsion. The location of the transition state on the reaction coordinate is very similar for the three substrates, at a difference of distances of around 0 Å. Of these, NVA's C^δ has the lowest abstraction barrier, perhaps because of the flexibility of the alkyl head. This is in accord with the experimental literature [23], which shows decay of the oxoferryl intermediate to be the fastest from this substrate, although the same experimental literature shows ABA to undergo this step more rapidly than THR, the opposite preference to that observed here. For all substrates, reaction was through the π channel, with $\text{Fe}-\text{O}-\text{H}$ angles of 96° , 100° and 92° for THR, ABA and NVA, respectively. In each of these cases, the structure of **TS_{DE}** was similar to the corresponding structure of **D**, but with an elongation of the $\text{Fe}-\text{O}_\text{P}$ bond by 0.12-0.17 Å, an increase of the $\text{N}_{\text{His235}}-\text{Fe}-\text{O}_\text{P}$ angle of $14-20^\circ$, and an elongation of the $\text{Fe}-\text{N}_{\text{His235}}$ bond by 0.14–0.18 Å (Table 5.7). **TS_{DE}** for THR in Conformation 1 is shown in Figure 5.17.

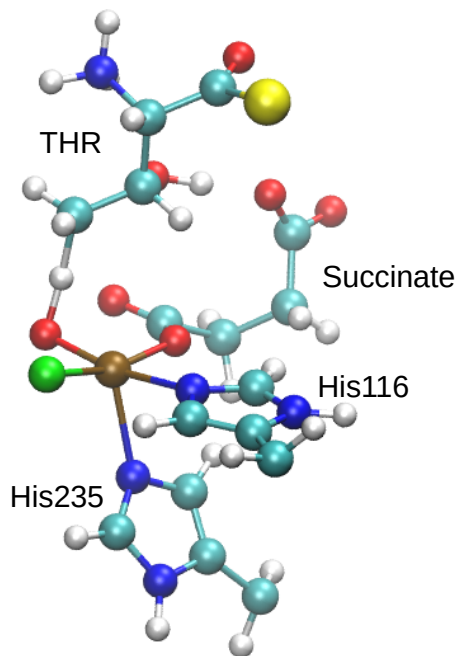


Figure 5.17: **TS_{DE}** for THR in Conformation 1

Whilst ABA's radical intermediate **E** is considerably higher in energy than those of the other two substrates, this is explained by the hydrogen bonding environment. For THR and NVA the newly formed ferric hydroxide group forms a hydrogen bond to the free succinate oxygen, whereas for ABA this hydroxide group faces in the other direction. Rotating the hydroxide to allow this interaction lowers the energy by 14 kJ/mol (see Figure 5.18).

The barrier to abstraction from NVA's C^γ was lower than that to abstraction from C^δ

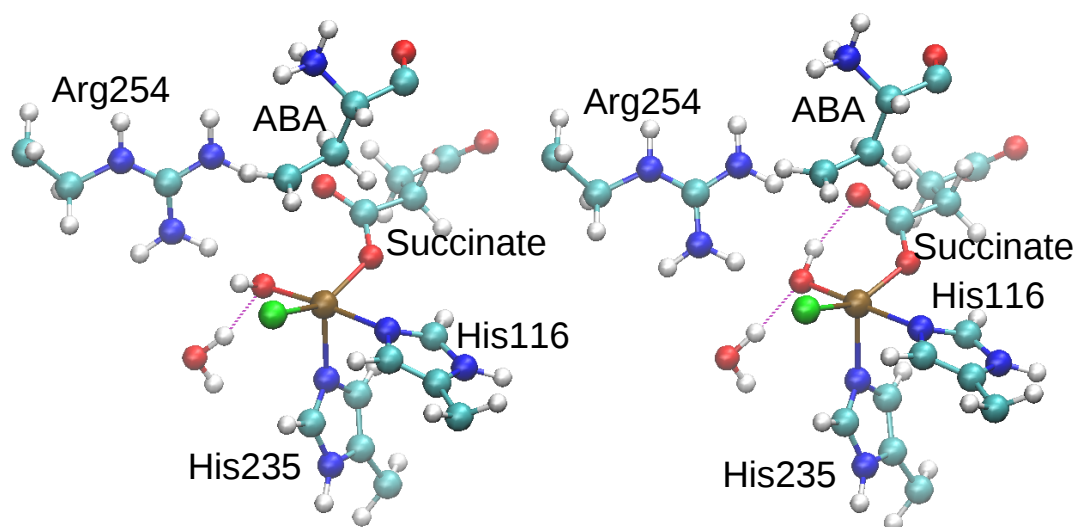


Figure 5.18: Rotational conformations of the ferric hydroxide group of **E** after hydrogen abstraction from ABA. Initially (left) this was hydrogen bonded only to a water molecule, but rotating to allow it to interact with succinate (right) lowered the energy by 14 kJ/mol

by almost 50 kJ/mol (Figure 5.27). However, the nature of the reaction was somewhat different. During abstraction from C^δ , the H–O distance was decreased primarily by the substrate head moving towards the oxoferryl oxygen, whereas during abstraction from C^γ the oxoferryl complex rearranged to bring the oxygen closer to the hydrogen. As a result, during abstraction from C^γ , the $N_{\text{His235}} - \text{Fe} - \text{O}_\text{P}$ angle is considerably higher for both **TS_{DE}** and **E** than for the corresponding species during abstraction from C^δ (Figure 5.19). For C^γ these angles are 174° and 151° for **TS_{DE}** and **E**, respectively, whereas for C^δ they are much lower at 116° and 114° , respectively). The significantly lower barrier from C^γ is probably a result of this geometric change—the Fe–O–H angle is consequently 115° for C^γ rather than 92° for C^δ , the former of which is far closer to the ideal angle for π -attack of around 120° . By moving backwards along the reaction coordinate from **TS_{DE}** for C^γ , and optimising freely, a new isomer of **D** is obtained, which retains the new orientation of O_P (with an N–Fe–O angle of 172° , corresponding to **D2**). This is 6 kJ/mol higher in energy than the structure of **D1** from which the scans started. Interestingly, for C^γ both **TS_{DE}** and the new structure of **D** have a different orientation of the succinate head, with the free carboxylate oxygen reoriented away from Arg254 (see Figure 5.19). The radical intermediate **E** for abstraction from C^γ is 16 kJ/mol lower in energy than the corresponding species for abstraction from C^δ . This difference is likely due to the fact the C^γ is a secondary site, providing stabilising inductive effects, but due to the different coordination geometry of the iron centre (N–Fe–O of 151° for C^γ vs 114° for C^δ) this

stabilising effect is certainly not the only factor in the energy separation of these two species.

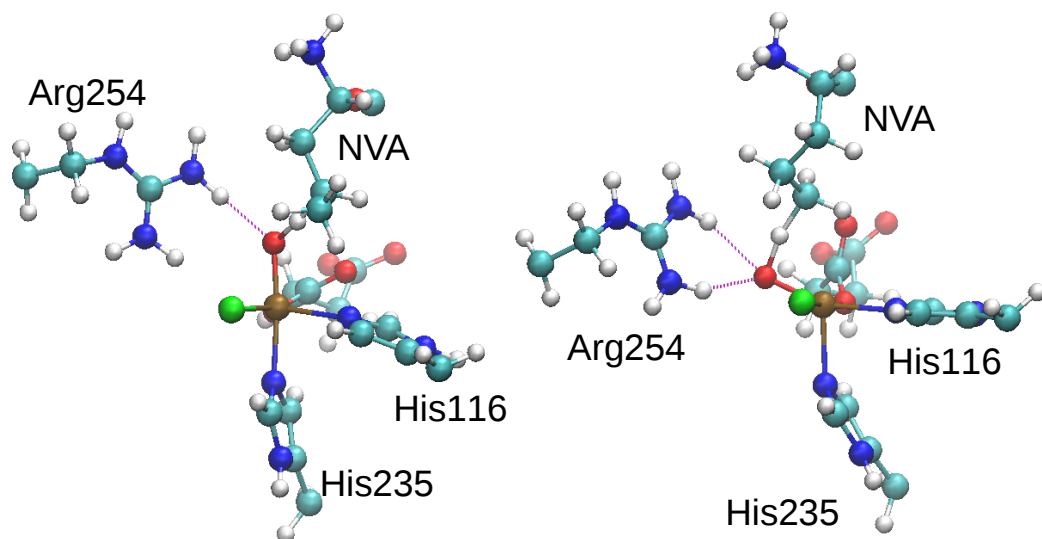


Figure 5.19: **TS_{DE}** for NVA C γ (left) and C δ (right). The former has a considerably higher N_{His235}–Fe–O_P angle (174° for C γ , compared to 116° degrees for C δ), as the substrate is less flexible at this position

Although **D** has a higher energy in THR's Conformation 2 than Conformation 1, the transition state and barrier are somewhat lower in the former conformation (Figure 5.27), by 10 and 24 kJ/mol respectively. This also undergoes π -channel reactivity, with an Fe–O–H angle of 95°. Like ABA, THR's Conformation 2 yielded a radical intermediate in which the ferric hydroxide did not have a hydrogen bond to succinate, leading to a higher energy. Rotating and reoptimising brought the radical intermediate's energy down by 32 kJ/mol. In both conformations, during optimisation, reversing the reaction coordinate from the transition state and freely optimising led to a lower energy conformer of **D**, due to relaxation of nearby protein residues.

In contrast to the results of the oxygen activation section, the results of this section show significant discrepancies from experiments. NVA, which has been shown by experiment to exhibit a preference for reaction from C δ , was calculated here to have a lower barrier to abstraction from C γ by 56 kJ/mol. THR, which is observed experimentally to have a lower barrier to abstraction than the other two substrates, is observed here, when in Conformation 2, to have the lowest barrier to abstraction of any substrate at the terminal methyl group. A possible explanation for this deviation from literature is the potential for

Hydrogen Abstraction

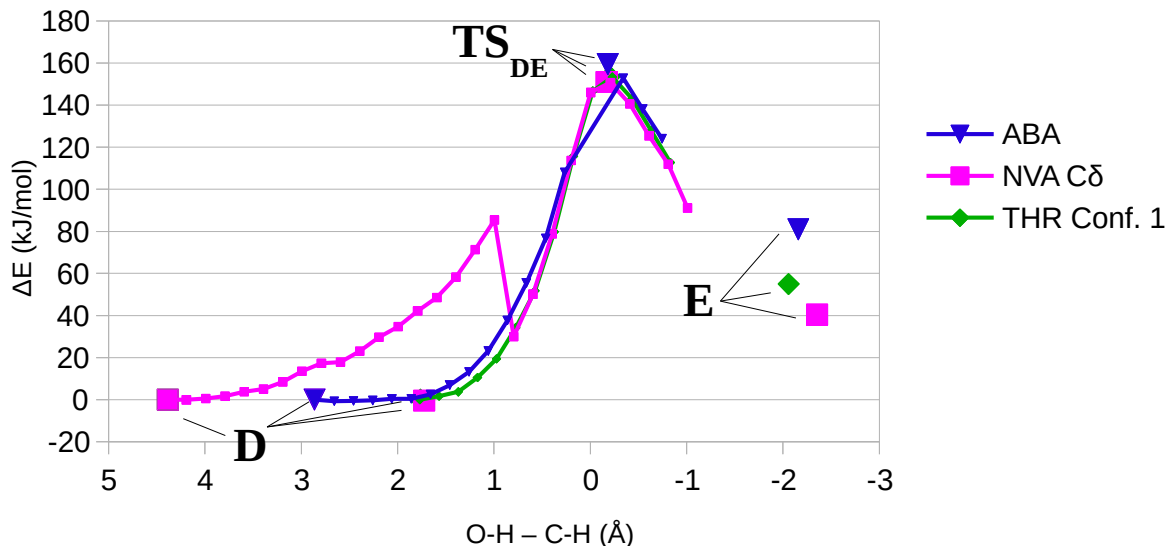


Figure 5.20: Energy profile of hydrogen abstraction from the three substrates. A secondary minimum for **D** (NVA C δ) was observed, corresponding to a rotation of the N – C $^{\alpha}$ – C $^{\beta}$ – C $^{\gamma}$ torsion, with an energy difference of 0.45 kJ/mol. Stationary points (optimised without restraints) are shown with larger symbols. All points plotted are at the def2-SVP basis level

interconversion between **D1** and **D2**. Both are observed here within the reaction profile of NVA, with an energy separation of only 6 kJ/mol in favour of **D1**, but a difference in the barriers to reaction of 56 kJ/mol in favour of **D2**, albeit from a different site. Investigation of the mechanism of this isomerism, and its implication to the reactions of other substrates, would certainly warrant further study.

These barriers are considerably higher than those observed in the computational literature. Wong and co-workers find abstraction barriers of 111 and 120 kJ/mol from **D1** and **D2** respectively, whilst Borowski and co-workers find abstraction barriers as low as 77 and 73 kJ/mol for **D3** and **D4** respectively. This may be due to the steric constraints imposed by the full protein environment of this QM/MM study, as both Wong and co-workers and Borowski and co-workers use cluster models. These steric constraints may be preventing the substrate from forming a favourable Fe–O–H angle.

5.3.3 Radical Rebound

Radical rebound was again controlled by a difference of distances, here the difference of C–Cl and Fe–Cl distances for chloride rebound and C–O and Fe–O distances for hydroxide rebound. Scans were run from all of the radical species formed in Section 5.3.2: ABA, THR in Conformations 1 and 2, and NVA at both C $^{\gamma}$ and C $^{\delta}$. In the two cases in which it was relevant (ABA and THR in Conformation 2), scans for both rotamers were run, but unless otherwise specified the following discussion refers to the conformation of **E** in which the ferric hydroxide group had a hydrogen bond to the nearby succinate oxygen. In many cases the reaction profiles led to an unexpected result—two pathways to radical rebound were feasible.

Figure 5.21 shows chloride rebound to ABA, THR Conformation 1, and NVA C $^{\delta}$. The scan for ABA is noticeably different from the other two cases, being much smoother and lower. It yielded a transition state slightly below the scan curve, as a result of a torsion of the succinate backbone. Reoptimising to **E** with this torsion lowered the energy by 9 kJ/mol. For THR and NVA, the energy rises significantly higher, then falls sharply. Reversing the direction of the scan here led to an entirely different pathway (see Figure 5.23).

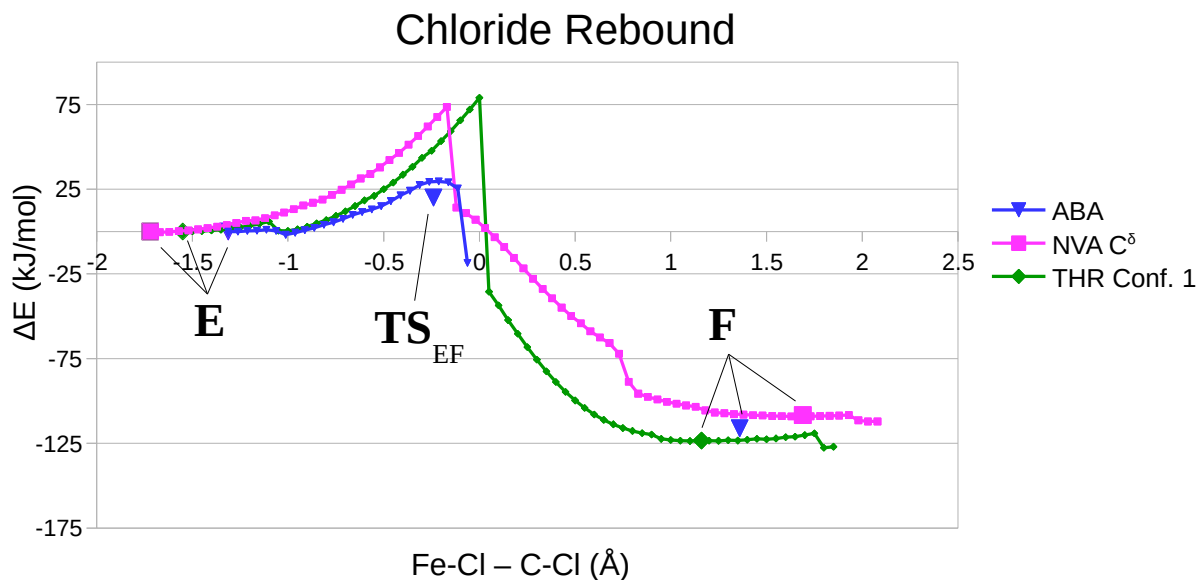


Figure 5.21: Energy profile of chloride rebound. All points plotted are at the def2-SVP level. Larger symbols represent stationary points.

The backward scans yielded a reactant that was a variant structure of **E** (here denoted

E2) in which the substrate radical is coordinated to the Fe centre (Figure 5.22). This species retains the unpaired electron on the carbon centre, evidenced by this carbon’s spin density ($\rho_u \approx -0.7$), but has a significant degree of iron-radical interaction, evidenced by the partial pyramidalisation of the radical methylene carbon. This pathway leads to a lower energy transition state in absolute terms, but a significantly higher reaction barrier—this was only 3 kJ/mol for ABA from **E**, but 55 and 78 kJ/mol for THR and NVA respectively from **E2**.

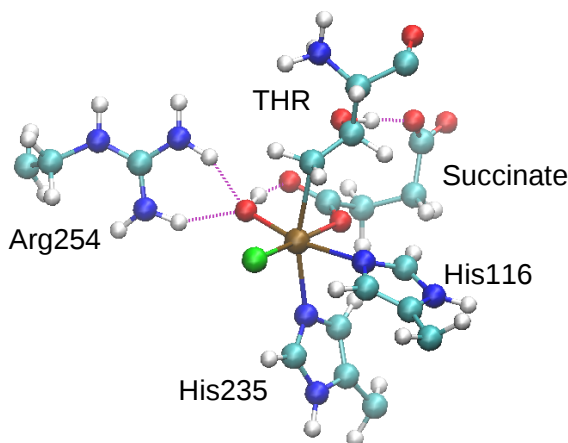


Figure 5.22: Species **E2** formed during Cl rebound with THR. Note the iron-radical interaction.

Structurally, **TS_{EF}** for ABA had an only slightly increased $N_{\text{His235}} - \text{Fe} - \text{Cl}$ angle (128° for **TS_{EF}** vs 105° for **E**), with the reactive carbon and chlorine brought together largely by motion of the carbon (Figure 5.24). Reversing the scan coordinate from the transition state and freely optimising led to a lower-energy conformer of the succinate tail.

The equivalent scans for hydroxide rebound are shown in Figure 5.25. The scans of THR and ABA yielded a structure for **TS_{EF'}**. The scans for both ABA and NVA yielded secondary minima of **E**, as a result of a torsion of the succinate backbone. These structures of **E** both had lower energies than the starting point. In addition, for ABA, optimisation of **TS_{EF}** led to an even more stable arrangement of the succinate tail and environment, which, upon reversal of the scan coordinate, led to a conformer of **E** lower in energy by a full 32 kJ/mol than that from which the scan started. The scan for NVA again fell onto the **E2–F'** pathway—reversing the direction of the reaction coordinate for this scan yielded **E2**. Unlike **TS_{EF}**, the distance between the carbon radical and its rebound partner during formation of **TS_{EF'}** was in both cases decreased partly by a change in the $N_{\text{His235}} - \text{Fe} - \text{O}_\text{P}$ angle, as well as the movement of the carbon towards the oxygen (see Figure 5.24 and

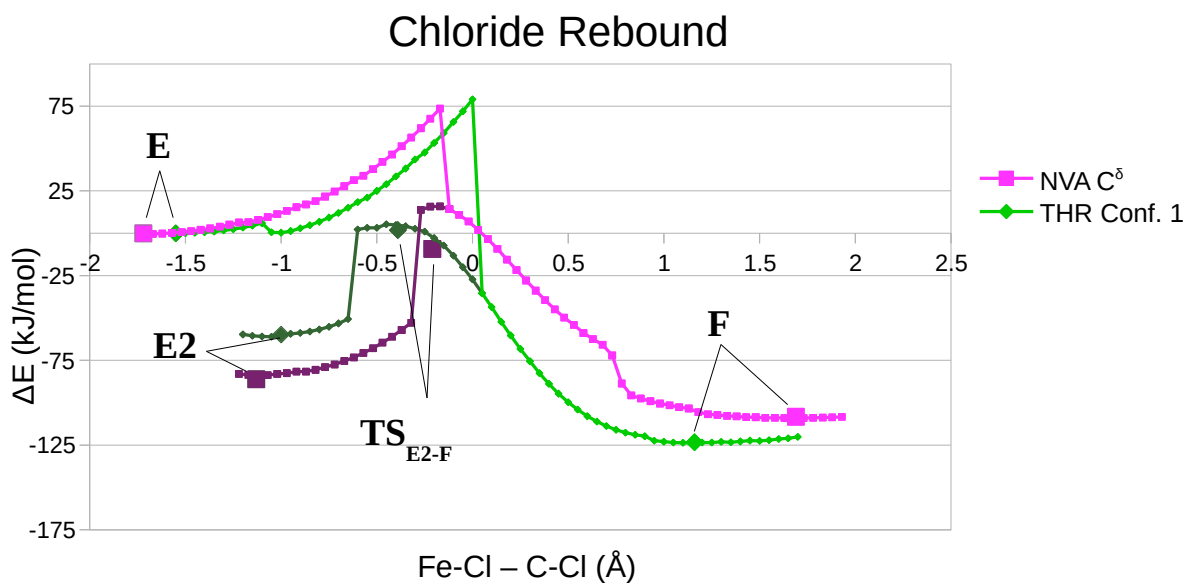


Figure 5.23: Energy profile of chloride rebound for NVA C^δ and THR Configuration 1, with reverse scans included. All plotted points are at def2-SVP level. Larger points correspond to stationary points, darker points to the reversed scan direction

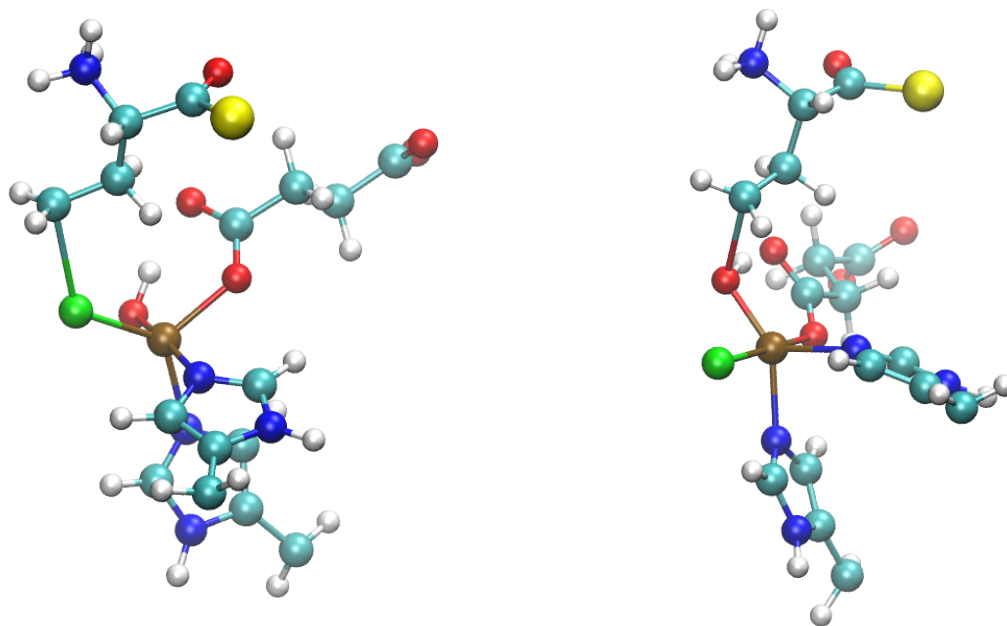


Figure 5.24: TS_{EF} (left) and $TS_{EF'}$ (right) for the substrate ABA

Table 5.7) In both cases this had an angle of around 140° , and corresponded to an increase of around 40° .

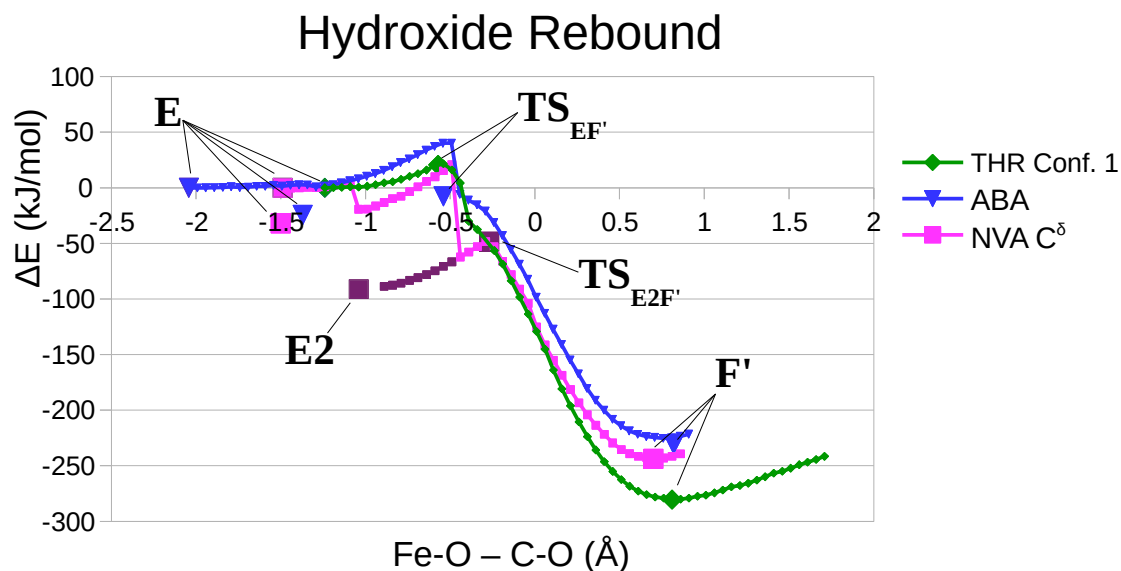


Figure 5.25: Energy profile of hydroxide rebound. Larger symbols represent stationary points, whilst darker colours represent scan points in the opposite direction (in the latter case only applying to NVA C^δ). Note that NVA C^δ and ABA both yield a more stable conformation of **E** as a secondary minimum during the scan.

5.3.4 Summary

Overview

Figure 5.26 shows all available stationary points for ABA, THR in Conformation 1, and NVA C^δ . From these data, it is evident that hydroxide rebound yields a significantly lower energy product than chloride rebound. However, it is also clear that both of these reactions are irreversible, meaning that the selectivity cannot be controlled by relative product stability. ABA, the only substrate to for which both TS_{EF} and $TS_{EF'}$ were obtained, had fairly low barriers to either reaction, in accord with its reactivity.

Figure 5.27 compares all available stationary points from THR in Conformations 1 and 2 to NVA reacting at both C^δ and C^γ . Again, chlorinated products are uniformly less stable than hydroxylated ones, but, the reactions being irreversible, selectivity must be controlled by the transition state energies. The low barrier to hydroxylation to THR in Conformation 1 suggests that THR reacts from Conformation 2 (as this reaction is never observed experimentally), which is supported by the considerably lower barrier to hydrogen abstraction from THR in this conformation. It is worth noting that at no point did proton transfer between the substrate ammonium group and the carboxylate group of Glu102 occur for any substrate. This is significant, as docking and molecular dynamics simulations were all carried out on the assumption that the protonation states of these

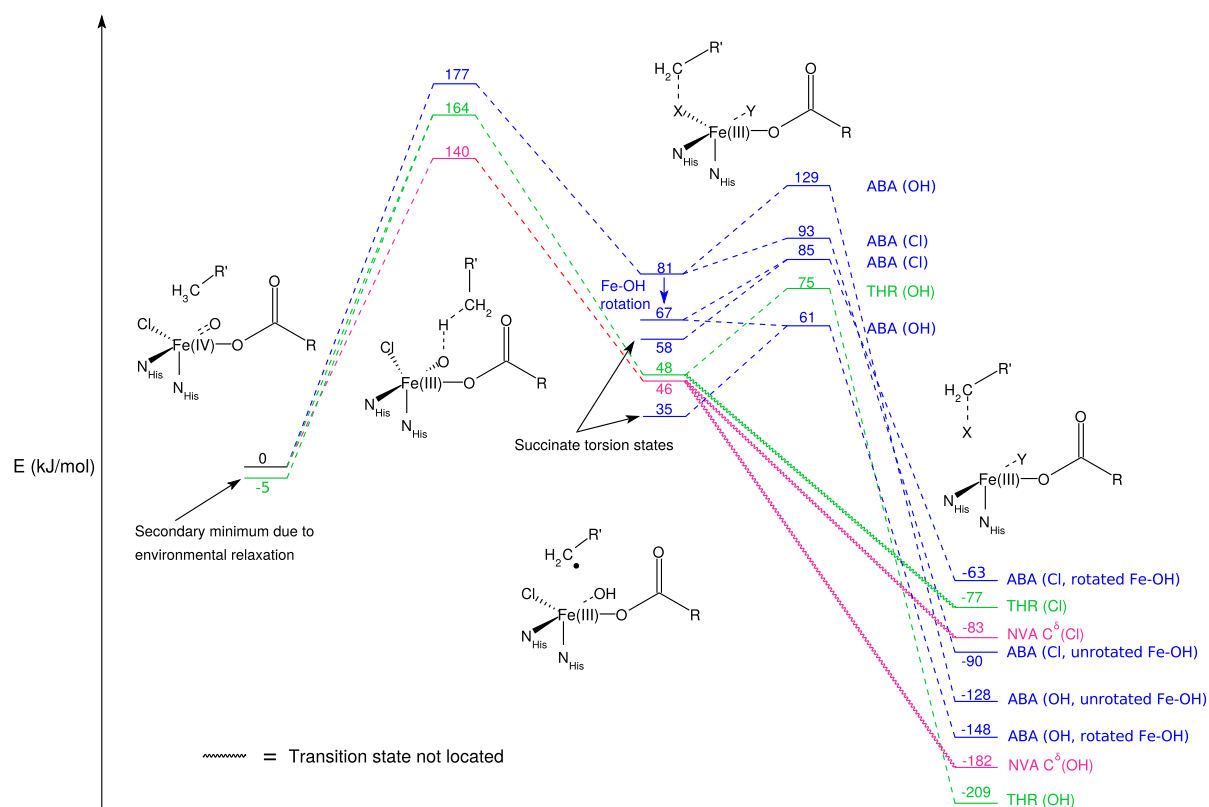


Figure 5.26: Energy profile of hydrogen abstraction and radical rebound for ABA, THR Conf. 1 and NVA C^δ .

two groups were fixed.

THR

From the structure of **B** with bound THR, the oxygen adduct **C** was observed to have a septet ground state, which has been observed before in the literature for FT-NHFe enzymes [112]. Fe-H distances showed excellent agreement with those obtained experimentally by Martinie and co-workers, to within 0.1 Å. Although $^3\mathbf{C}$ was observed to have a different oxidation state to all other structures of **C** (for any substrate), with a ferrous iron bound to a neutral oxygen diradical rather than a ferric iron with peroxide ligand, oxygen activation scans from all multiplicities of **C**, including this aberrant triplet, led to a similar structure of **D1**.

From **D1**, a structure was prepared for Conformation 2, which led to a fork in the subsequent reaction scans. **D** in Conformation 2 was observed to have a higher energy than Conformation 1 by 18 kJ/mol, in accord with its lower occupancy during MD simulations. Hydrogen abstraction from Conformation 2 led to a barrier lower by 24 kJ/mol, but also

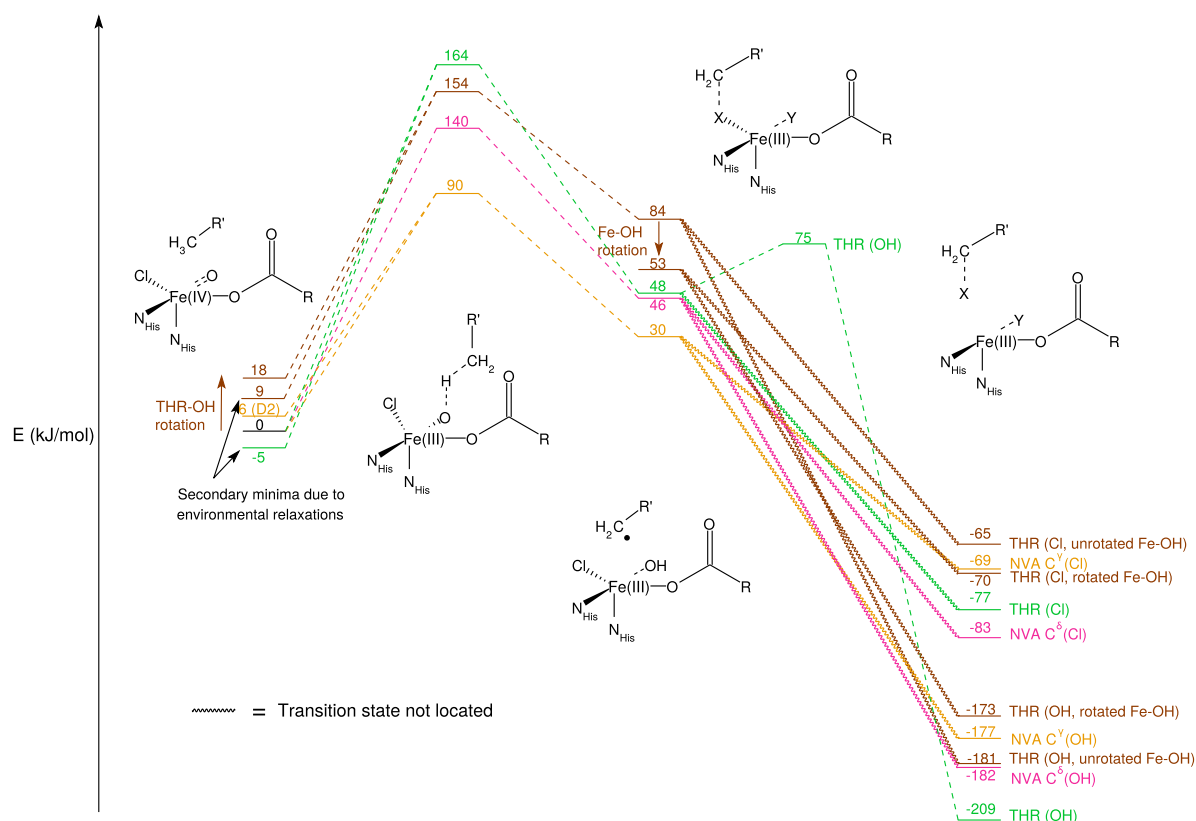


Figure 5.27: Energy profile of hydrogen abstraction and radical rebound for THR Confs. 1 and 2 and NVA C^γ and C^δ.

to a product **E** with an unfavourable orientation of the ferric hydroxide group. Rotation of this group to allow a hydrogen bond to the adjacent succinate oxygen lowered the energy by 31 kJ/mol, although the absolute energy of this species was still 5 kJ/mol higher than that of **E** in Conformation 1.

Of the six subsequent radical rebound scans (chloride and hydroxide rebound to THR in each of Conformation 1 and Conformation 2, the latter both with and without the ferric hydroxide–succinate hydrogen bond), the only transition state obtained from **E** was **TS_{EF}** for THR in Conformation 1. As this has a low barrier, of 27 kJ/mol, whilst this reaction is never observed experimentally, it would suggest that THR reacts from Conformation 2.

ABA

Adduct formation from the structure of **B** with bound ABA also yielded structures of **C** in all three of the triplet, quintet and septet states, again with a septet ground state. Here, however, all three had a similar electronic structure, with ferric iron bound to a peroxide ligand. Again, the Fe–H distances were in excellent agreement with those reported from

the experiments of Martinie and co-workers, all falling within 0.2 Å of the experimental results. Oxygen abstraction from ^5C and ^7C again led to **D1**, although on the triplet surface this led to **D2**. As the scan from ^3C had a higher energy for the reactant, product and scan curve than on the other two surfaces, **D1** was selected for subsequent hydrogen abstraction scans.

Hydrogen abstraction again led to a rotamer of the ferric hydroxide group that did not have a hydrogen bond to the succinate oxygen. Rotating this hydroxide group to allow this interaction lowered the energy by 14 kJ/mol. Subsequent radical rebound scans yielded all four possible radical rebound transition states (TS_{EF} and $\text{TS}_{\text{EF}'}$ from **E** with both rotamers of the ferric hydroxide group). Reversing the scan coordinate from either of the transition states from the rotamer with hydroxide–succinate interaction led to a structure of **E** which, due to a torsional change in the succinate tail, was more favourable by 9 kJ/mol. Barriers to radical rebound from both rotamers of **E** were all lower than 50 kJ/mol, with the rotamer without hydroxide–succinate interaction showing a preference for chlorination, and the other rotamer showing the (experimentally observed) opposite preference.

NVA

Adduct formation from the structure of **B** with bound NVA led to some surprising observations. In the septet state, which was the ground state for the other two substrates, **C** was unstable, with the oxygen not binding to the iron complex. The quintet state, which was the only multiplicity to facilitate oxygen binding, had a different electronic structure to that of ^5C for the other two substrates, as it exhibited antiferromagnetic coupling between the ferric iron and its peroxide ligand. Oxygen activation again led to **D1**, which had very similar spin populations to the other two substrates.

Hydrogen abstraction from C^δ was similar to that from the other two substrates, although it followed a change in the $\text{N} - \text{C}^\alpha - \text{C}^\beta - \text{C}^\gamma$ torsion from the conformation present during oxygen activation. Abstraction from C^γ followed isomerism from **D1** to **D2**. This had a considerably lower barrier, suggesting a preference for hydrogen abstraction from this isomer of **D** (probably due to the more favourable Fe–O–H angles it allows). Unfortunately, radical rebound scans did not lead to a structure of TS_{EF} or $\text{TS}_{\text{EF}'}$ for either reactive carbon.

	Parameter	B	C	D	TS_{DE}	E	TS_{EF}	F	TS_{EF'}	F'
THR (Conf. 1)	Fe–N _{His235} (Å)	2.09	1.99	2.10	2.26	2.15	—	2.17	2.17	2.17
	Fe–N _{His116} (Å)	2.10	2.02	2.10	2.08	2.05	—	2.12	2.11	2.16
	Fe–Cl (Å)	2.33	2.31	2.25	2.26	2.25	—	n/a	2.27	2.32
	Fe–O _P (Å)	n/a	1.95	1.63	1.75	1.83	—	1.92	1.87	n/a
	Fe–O _K (Å)	2.34	2.29	1.93	1.92	1.97	—	2.07	1.97	2.04
	N _{His235} –Fe–O _P (°)	n/a	164	95	114	107	—	101	140	n/a
THR (Conf. 2)	Fe–N _{His235} (Å)	2.09	1.99	2.11	2.25	2.16	—	2.16	—	2.16
	Fe–N _{His116} (Å)	2.10	2.02	2.10	2.08	2.04	—	2.12	—	2.09
	Fe–Cl (Å)	2.33	2.31	2.25	2.24	2.24	—	n/a	—	2.35
	Fe–O _P (Å)	n/a	1.95	1.64	1.78	1.84	—	1.92	—	n/a
	Fe–O _K (Å)	2.34	2.29	1.93	1.93	1.96	—	2.08	—	2.13
	N _{His235} –Fe–O _P (°)	n/a	164	100	118	114	—	113	—	n/a
ABA	Fe–N _{His235} (Å)	2.11	2.00	2.07	2.25	2.08	2.14	2.12	2.13	2.17
	Fe–N _{His116} (Å)	2.10	2.01	2.10	2.04	2.02	2.05	2.11	2.10	2.12
	Fe–Cl (Å)	2.31	2.29	2.26	2.26	2.25	2.29	n/a	2.35	2.34
	Fe–O _P (Å)	n/a	1.94	1.64	1.76	1.83	1.88	1.91	1.89	n/a
	Fe–O _K (Å)	2.35	2.32	1.94	1.93	2.00	2.05	2.06	2.13	2.08
	N _{His235} –Fe–O _P (°)	n/a	171	93	107	91	90	99	143	n/a
NVA (C ^δ)	Fe–N _{His235} (Å)	2.10	2.12	2.05	2.19	2.15	—	2.11	—	2.10
	Fe–N _{His116} (Å)	2.11	2.13	2.12	2.06	2.04	—	2.14	—	2.09
	Fe–Cl (Å)	2.36	2.28	2.29	2.27	2.29	—	n/a	—	2.36
	Fe–O _P (Å)	n/a	2.04	1.63	1.80	1.81	—	1.93	—	n/a
	Fe–O _K (Å)	2.43	2.25	1.95	1.88	1.99	—	2.05	—	2.26
	N _{His235} –Fe–O _P (°)	n/a	158	96	116	114	—	116	—	n/a
NVA (C ^γ)	Fe–N _{His235} (Å)	2.10	2.12	2.10	2.05	2.03	—	2.16	—	2.14
	Fe–N _{His116} (Å)	2.11	2.13	2.11	2.12	2.13	—	2.10	—	2.12
	Fe–Cl (Å)	2.36	2.28	2.31	2.32	2.31	—	n/a	—	2.40
	Fe–O _P (Å)	n/a	2.04	1.63	1.75	1.85	—	1.91	—	n/a
	Fe–O _K (Å)	2.43	2.25	1.94	1.94	1.90	—	2.02	—	1.99
	N _{His235} –Fe–O _P (°)	n/a	158	172	174	151	—	115	—	n/a

Table 5.7: Geometric parameters of species **B** to **F'** on the quintet surface. For species **E** to **F'** for ABA and THR Conf. 2, the parameters listed are for the lower energy rotamer of the ferric OH group. Where multiple minor conformers are possible (**D** for THR Confs. 1 and 2 and **E** for ABA), values listed are for the conformation from which the scan commenced

5.4 Conclusions

The labyrinthine pathways for reactivity that were identified, across multiple spin-surfaces and hydrogen-bonding configurations, make for an intriguing picture of SyrB2’s mechanism. The presence of multiple low-lying excited spin-states of the oxygen adduct species **C** was confirmed, with the septet identified as the ground state where all three spin-states were available. In addition, multiple electronic configurations for each multiplicity were identified, arising from the potential for either ferromagnetic or antiferromagnetic coupling between electrons on the iron and the bound dioxygen. Due to the different electronic structures of **C** observed in the presence of different substrates, it appears that the electronic structure of this species is minutely sensitive to changes in its geometry.

During oxygen activation, with a single exception, all available permutations of the three substrates and the three multiplicities yielded the isomer **D1**. Subsequently, during hydrogen abstraction, the barriers are fairly similar for THR Conformation 1 and ABA, but considerably lower for NVA’s C^δ . Interestingly, NVA’s C^γ has a significantly lower hydrogen abstraction barrier than C^δ , probably because a change in isomerism led to an angle of attack that lends itself well to π -channel reactivity. THR’s Conformation 2 yielded a lower hydrogen abstraction barrier than Conformation 1, which may go some way to explaining the preference for chlorination later in the reaction—Conformation 1 was shown to have a very low barrier to hydroxylation, a reaction never observed experimentally for this substrate, so it is unlikely that THR reacts from this conformation.

Finally, an alternative pathway to radical rebound was identified, which proceeds from the Fe-bound radical **E2**. Formation of this species leads to very high barriers to chlorination and hydroxylation, as it is far more stable than **E**. Unfortunately, transition states in this pathway are considerably lower in absolute energy than the corresponding transition states for reaction from **E**, which led to difficulties in finding these transition states from the latter pathway. Although the hydroxylated products are generally considerably lower in energy than the chlorinated ones, since both reactions are irreversible selectivity must be determined by the barrier heights.

Chapter 6

Conclusions

Whilst this study set out to answer specific questions about SyrB2’s reactivity, this endeavour has been indivisible from an investigation into reliable models and techniques. The intricate sets of hydrogen bonding networks, able to hold nearly identical substrates in very different positions, have added dimensions of complexity to a reaction already marked by the presence of high-spin ferric and ferryl species with numerous low-lying excited states.

The docking simulations found essentially identical docked poses for the three substrates, dominated by a salt-bridge between the substrate’s ammonium group and the carboxylate group of Glu102. Hydrophobic interactions towards the substrates pantetheine tails and alkyl heads also stabilise the pose, with the influence of THR’s γ -hydroxide group and NVA’s additional methylene unit having little to no effect on the overall position of the substrate.

When the protein-substrate complexes are more free to move, however, entirely unanticipated effects come into play. The tail of 2-oxoglutarate is willing to disrupt an intricate network of hydrogen bonds involving Ser237, Thr113, Arg248 and Trp145 to form a hydrogen bond to the γ -hydroxide group of THR. This interaction is persistent, leading to a new, stable position for THR marked by the cleavage of the direct salt bridge between its ammonium group and Glu102, and its replacement by a longer-distance, water-mediated interaction. This leads to a radically different position of THR in comparison to the other two substrates.

When oxygen adducts were formed in the presence of the equilibrated structures of THR and ABA, the distances between Fe and the hydrogen atoms of the terminal carbon matched those from experimental data from literature almost exactly. This was not the case for NVA—the choice of the torsional state of the substrate head, driven by the need to clear space for the dioxygen to bind, placed the terminal methyl group some distance from the Fe centre. However, as the distances measured by Martinie and co-workers came

from experiments run on an NO adduct species more stable and long-lived than **C**, it is quite possible that oxygen binds to the Fe complex preferentially when NVA has a more open conformation, but that if the system is given time to equilibrate to **C** NVA will adopt a more closed conformation.

The electronic structure of species **C** is complicated. Whilst the septet is the ground state in each case for which it is stable, low-lying quintet and triplet states were both observed, with multiple electronic configurations made possible through ferromagnetic or antiferromagnetic coupling. These findings add to the general discord within the literature, which often identifies different ground states for complexes of type **C** or **AC** even with the same functional or ligand environment.

Oxygen activation followed a similar mechanism on most spin surfaces. Generally, a single intermediate (**G** for scans on the quintet or triplet surfaces, or **J** on the septet surface) preceded a highly exothermic decarboxylation step and the formation of a bridged intermediate. After cleaving of the O–O bond, with the exception of the scan carried out on the triplet surface in the presence of ABA, all scans formed the isomer **D1**.

These structures of **D1** were used for subsequent QM/MM calculations. MD simulations of ABA and NVA in the presence of this isomer found these two substrates not to move significantly from their docked conformations. THR, however, was shown to adopt a second state, facilitated by a new hydrogen bonding partner of its γ -hydroxide group. A second structure was prepared to represent this during QM/MM calculations. The possibility of reaction from other isomers was not investigated in great detail at the QM/MM level principally because all but one of the oxygen-activation scans led to **D1**. This would, however, make for an interesting line of inquiry for follow-up studies.

Method validation for DFT was carried out on models of species **D**. All of the density functionals tested, as well as all coupled-cluster calculations, correctly identified the quintet as the ground state of this species. Whilst the D_1 diagnostics of these coupled cluster-calculations strongly suggested that they were not of benchmark quality, the use of KS reference orbitals brought the values to a reasonably low level. Based on the consensus between RKS/CCSD(T) and CASPT2 from literature, the B3LYP functional was selected for QM/MM calculations.

NVA had a lower barrier to hydrogen abstraction than the other two substrates, from both C^γ and C^δ , which is in accord with the experimentally determined faster rate of decay of the oxoferryl species in the presence of this substrate. Interestingly, C^γ had a lower barrier to hydrogen abstraction than C^δ , possibly because isomerism to **D2** led to a different Fe – O – H angle. The possibility of isomerism to **D2** in the presence of the other substrates certainly warrants further investigation. THR experienced a significantly (24 kJ/mol) lower barrier to hydrogen abstraction from Conformation 2 than from Con-

formation 1. ABA, however, had a somewhat higher barrier than the other two substrates. This is not in agreement with the experimentally determined rates of decay of the oxoferryl species, which place ABA between NVA and THR.

Attempts to model radical rebound led to a second structure of the radical intermediate. This structure, **E2**, in which the substrate radical is bound to the Fe-centre, is considerably more stable than the free alkyl radical species **E**. Radical rebound from **E2** led to a significantly higher barrier than that from **E**, but to a transition state considerably lower in absolute energy. The barrier heights must determine selectivity, since whilst the chlorinated products uniformly have a higher energy than the hydroxylated ones, both reactions appear to be irreversible. As such, selectivity cannot conclusively be explained without a full set of transition states for the reaction **E**–**F**. It is the author’s profound hope that future studies can provide these, elucidate the roles of **D2** and **E2**, and ultimately solve the perplexing problem of SyrB2’s reactivity, and that the findings of this thesis are of use to them in so-doing.

Appendix A

Supplementary Data from QM/MM Calculations

A.1 Supplementary data at the def2-TZVP Basis Level

Parameter	THR (Conf.1)			ABA		NVA (C ^δ)	
	⁷ J	E2	TS _{E2F}	⁵ I	⁷ J	E2	TS _{E2F}
Fe–N _{His235} (Å)	2.09	2.23	2.26	2.10	2.11	2.18	2.09
Fe–N _{His116} (Å)	2.11	2.19	2.13	2.10	2.17	2.19	2.22
Fe–Cl (Å)	2.25	2.32	2.29	2.33	2.24	2.34	2.35
Fe–O _P (Å)	1.94	1.85	1.89	1.77	1.98	1.88	1.92
Fe–O _k (Å)	2.29	2.07	1.96	2.17	2.21	2.10	2.08
N _{His235} –Fe–O _P (°)	155	113	86	127	166	109	118

Table A.1: Geometric parameters of various species at the def2-TZVP level

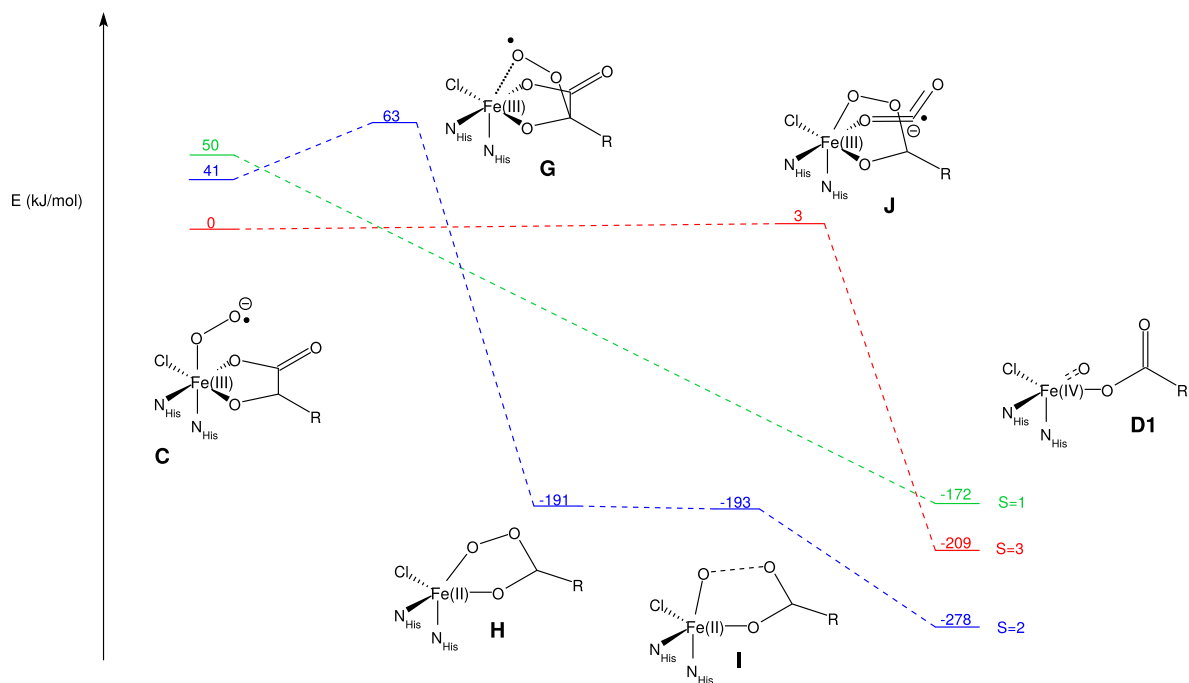


Figure A.1: Energy profile of oxygen activation for ABA at the def2-TZVP basis level

A.2 Geometries at the def2-SVP Basis Level

Parameter	THR			ABA			NVA
	S=1	S=2	S=3	S=1	S=2	S=3	S=2
Fe-N _{His235} (Å)	2.02	2.00	2.07	1.97	2.00	2.09	2.11
Fe-N _{His116} (Å)	2.11	2.02	2.12	2.00	2.05	2.11	2.13
Fe-Cl (Å)	2.32	2.31	2.27	2.25	2.29	2.25	2.29
Fe-O _P (Å)	2.01	1.97	2.20	1.95	1.95	2.14	2.06
Fe-O _{A1} (Å)	2.06	1.97	2.04	1.94	1.99	2.04	2.01
Fe-O _K (Å)	2.30	2.29	2.12	1.99	2.31	2.29	2.25
O-O (Å)	1.24	1.28	1.27	1.28	1.28	1.27	1.27
N _{His235} -Fe-O _P (°)	169	163	155	178	170	162	156

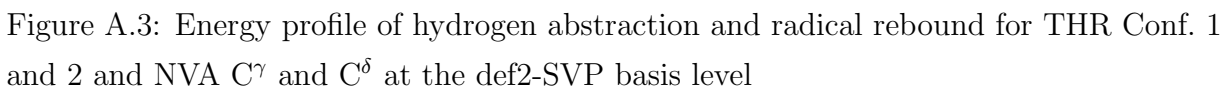
Table A.2: Geometric parameters of the oxygen adduct **C** at the def2-SVP basis level

	Parameter	B	C	D	TS_{DE}	E	TS_{EF}	F	TS_{EF'}	F'
THR (Conf. 1)	Fe-N _{His235} (Å)	2.10	2.00	2.11	2.23	2.17	—	2.19	2.18	2.17
	Fe-N _{His116} (Å)	2.11	2.02	2.09	2.07	2.06	—	2.13	2.10	2.15
	Fe-Cl (Å)	2.34	2.31	2.26	2.26	2.26	—	n/a	2.29	2.32
	Fe-O _P (Å)	n/a	1.97	1.62	1.75	1.81	—	1.90	1.86	n/a
	Fe-O _K (Å)	2.26	2.29	1.92	1.93	1.97	—	2.07	1.99	2.08
	N _{His235} -Fe-O _P (°)	n/a	163	95	79	110	—	99	136	n/a
THR (Conf. 2)	Fe-N _{His235} (Å)	2.10	2.00	2.13	2.22	2.16	—	2.16	—	
	Fe-N _{His116} (Å)	2.11	2.02	2.09	2.08	2.05	—	2.14	—	
	Fe-Cl (Å)	2.34	2.31	2.26	2.26	2.25	—	n/a	—	
	Fe-O _P (Å)	n/a	1.97	1.63	1.78	1.84	—	1.92	—	
	Fe-O _K (Å)	2.26	2.29	1.93	1.93	1.96	—	2.08	—	
	N _{His235} -Fe-O _P (°)	n/a	163	101	116	116	—	113	—	
ABA	Fe-N _{His235} (Å)	2.10	2.00	2.07	2.24	2.09	2.14	2.13	2.14	2.19
	Fe-N _{His116} (Å)	2.10	2.05	2.10	2.04	2.02	2.05	2.12	2.10	2.12
	Fe-Cl (Å)	2.30	2.29	2.27	2.28	2.26	2.31	n/a	2.35	2.34
	Fe-O _P (Å)	n/a	1.95	1.63	1.75	1.82	1.87	1.90	1.87	n/a
	Fe-O _K (Å)	2.33	2.31	1.94	1.94	2.00	2.04	2.06	2.14	2.10
	N _{His235} -Fe-O _P (°)	n/a	170	92	107	92	90.4	99	140	n/a
NVA (C ^δ)	Fe-N _{His235} (Å)	2.11	2.12	2.06	2.15	2.15	—	2.11	—	2.11
	Fe-N _{His116} (Å)	2.12	2.13	2.11	2.07	2.04	—	2.17	—	2.09
	Fe-Cl (Å)	2.36	2.29	2.30	2.31	2.30	—	n/a	—	2.35
	Fe-O _P (Å)	n/a	2.06	1.62	1.75	1.80	—	1.92	—	n/a
	Fe-O _K (Å)	2.39	2.25	1.96	1.94	1.99	—	2.05	—	2.25
	N _{His235} -Fe-O _P (°)	n/a	156	96	110	114	—	115	—	n/a
NVA (C ^γ)	Fe-N _{His235} (Å)	2.11	2.12	2.09	2.05	2.06	—	2.21	—	2.15
	Fe-N _{His116} (Å)	2.12	2.13	2.11	2.13	2.12	—	2.15	—	2.12
	Fe-Cl (Å)	2.36	2.29	2.31	2.32	2.31	—	n/a	—	2.42
	Fe-O _P (Å)	n/a	2.06	1.62	1.74	1.84	—	1.90	—	n/a
	Fe-O _K (Å)	2.39	2.25	1.95	1.95	1.90	—	2.17	—	2.01
	N _{His235} -Fe-O _P (°)	n/a	156	171	173	142	—	120	—	n/a

Table A.3: Geometric parameters of species **B** to **F'** on the quintet surface at the def2-SVP basis level. For ABA and THR Conf. 2 species **E–F'**, the parameters listed are for the lower energy rotamer of the ferric OH group. Where multiple minor conformations are possible (**D** for THR Confs. 1 and 2 and **E** for ABA), values listed are for the conformation from which the scan commenced

Parameter		THR			ABA			NVA
		<i>S</i> = 1	<i>S</i> = 2	<i>S</i> = 3	<i>S</i> = 1	<i>S</i> = 2	<i>S</i> = 3	<i>S</i> = 2
dzvp	ΔE (kJ/mol)	20	50	0	51	42	0	n/a
	$\rho_u(\text{Fe})$	3.39	2.91	4.17	1.08	2.94	4.13	4.14
	$\rho_u(\text{O}_\text{P})$	−0.73	0.38	0.65	0.50	0.37	0.69	−0.21
	$\rho_u(\text{O}_\text{D})$	−0.81	0.52	0.60	0.51	0.48	0.65	−0.46

Table A.4: Spin-state separations and spin populations of the oxoferryl complexes **D** at the def2-SVP basis level. *The ABA/*S*=1 scan ended with the isomer **D2**



Bibliography

- [1] Butler, A.; Sandy, M. *Nature* **2009**, *460*, 848–854.
- [2] Senn, H. M. *Front. Chem.* **2014**, *2*, Article 98.
- [3] Hager, L. P.; Morris, D. R.; Brown, F. S.; Eberwein, H. *J. Biol. Chem.* **1966**, *241*, 1769–1777.
- [4] Vilter, H. *Bot. Mar.* **1983**, *26*, 429–435.
- [5] Vilter, H. *Phytochemistry* **1984**, *23*, 1387 – 1390.
- [6] Dairi, T.; Nakano, T.; Aisaka, K.; Katsumata, R.; Hasegawa, M. *Biosci. Biotechnol. Biochem.* **1995**, *59*, 1099–1106.
- [7] van Pée, K.-H.; Patallo, E. P. *Appl. Microbiol. Biotechnol.* **2006**, *70*, 631–641.
- [8] Ni, X.; Hager, L. P. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12866–12871.
- [9] Schmidberger, J. W.; James, A. B.; Edwards, R.; Naismith, J. H.; O’Hagan, D. *Angew. Chem.* **2010**, *122*, 3728–3730.
- [10] Vaillancourt, F. H.; Yeh, E.; Vosburg, D. A.; O’Connor, S. E.; Walsh, C. T. *Nature* **2005**, *436*, 1191–1194.
- [11] Vaillancourt, F. H.; Yin, J.; Walsh, C. T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10111–10116.
- [12] Blasiak, L. C.; Vaillancourt, F. H.; Walsh, C. T.; Drennan, C. L. *Nature* **2006**, *440*, 368–371.
- [13] Ueki, M.; Galonić, D. P.; Vaillancourt, F. H.; Garneau-Tsodikova, S.; Yeh, E.; Vosburg, D. A.; Schroeder, F. C.; Osada, H.; Walsh, C. T. *Chem. Biol.* **2006**, *13*, 1183–1191.
- [14] Khare, D.; Wang, B.; Gu, L.; Razelun, J.; Sherman, D. H.; Gerwick, W. H.; Hakansson, K.; Smith, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 14099–14104.

- [15] Galonic, D. P.; Vaillancourt, F. H.; Walsh, C. T. *J. Am. Chem. Soc.* **2006**, *128*, 3900–3901.
- [16] Fullone, M. R.; Paiardini, A.; Miele, R.; Marsango, S.; Gross, D. C.; Omura, S.; Ros-Herrera, E.; Bonaccorsi di Patti, M. C.; Laganà, A.; Pascarella, S.; Grgurina, I. *FEBS J.* **2012**, *279*, 4269–4282.
- [17] Pratter, S. M.; Ivkovic, J.; Birner-Gruenberger, R.; Breinbauer, R.; Zangger, K.; Straganz, G. D. *ChemBioChem* **2014**, *15*, 567–574.
- [18] Wong, C.; Fujimori, D. G.; Walsh, C. T.; Drennan, C. L. *J. Am. Chem. Soc.* **2009**, *131*, 4872–4879.
- [19] Hillwig, M. L.; Liu, X. *Nat. Chem. Biol.* **2014**, *10*, 921–923.
- [20] Kojima, T.; Leising, R. A.; Yan, S.; Que Jr, L. *J. Am. Chem. Soc.* **1993**, *115*, 11328–11335.
- [21] Purpero, V.; Moran, G. R. *J. Biol. Inorg. Chem.* **2007**, *12*, 587–601.
- [22] Vaillancourt, F. H.; Vosburg, D. A.; Walsh, C. T. *ChemBioChem* **2006**, *7*, 748–752.
- [23] Matthews, M. L.; Neumann, C. S.; Miles, L. A.; Grove, T. L.; Booker, S. J.; Krebs, C.; Walsh, C. T.; Bollinger, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 17723–17728.
- [24] Matthews, M. L.; Chang, W.-c.; Layne, A. P.; Miles, L. A.; Krebs, C.; Bollinger Jr, J. M. *Nat. Chem. Biol.* **2014**, *10*, 209–215.
- [25] Hanauske-Abel, H.; Günzler, V. *J. Theor. Biol.* **1982**, *94*, 421–455.
- [26] Wong, S. D.; Srnc, M.; Matthews, M. L.; Liu, L. V.; Kwak, Y.; Park, K.; Bell III, C. B.; Alp, E. E.; Zhao, J.; Yoda, Y.; Kitao, S.; Seto, M.; Krebs, C.; Bollinger, J. M.; Solomon, E. I. *Nature* **2013**, *499*, 320–323.
- [27] Matthews, M. L.; Krest, C. M.; Barr, E. W.; Vaillancourt, F. H.; Walsh, C. T.; Green, M. T.; Krebs, C.; Bollinger Jr, J. M. *Biochemistry* **2009**, *48*, 4331–4343.
- [28] Martinie, R. J.; Livada, J.; Chang, W.-c.; Green, M. T.; Krebs, C.; Bollinger, J. M.; Silakov, A. *J. Am. Chem. Soc.* **2015**, *137*, 6912–6919.
- [29] Pandian, S.; Vincent, M. A.; Hillier, I. H.; Burton, N. A. *Dalton Trans.* **2009**, 6201.

- [30] Kulik, H. J.; Blasiak, L. C.; Marzari, N.; Drennan, C. L. *J. Am. Chem. Soc.* **2009**, *131*, 14426–14433.
- [31] Borowski, T.; Noack, H.; Radoń, M.; Zych, K.; Siegbahn, P. E. M. *J. Am. Chem. Soc.* **2010**, *132*, 12887–12898.
- [32] Cramer, C. J. *Essentials of computational chemistry: theories and models*; J. Wiley: West Sussex, England ; New York, 2002.
- [33] Jensen, F. *Introduction to computational chemistry*, 2nd ed.; John Wiley & Sons: Hoboken NJ, 2006.
- [34] Harvey, J. N. *J. Biol. Inorg. Chem.* **2011**, *16*, 831–839.
- [35] Klopper, W.; Bachorz, R. A.; Hättig, C.; Tew, D. P. *Theor. Chem. Acc.* **2010**, *126*, 289–304.
- [36] Chen, H.; Lai, W.; Shaik, S. *J. Phys. Chem. Lett.* **2010**, *1*, 1533–1540.
- [37] Ye, S.; Riplinger, C.; Hansen, A.; Krebs, C.; Bollinger, J. M.; Neese, F. *Chem. Eur. J.* **2012**, *18*, 6555–6567.
- [38] Lee, T.; Taylor, P. *Int. J. Quant. Chem.* **1989**, *36*, 199–207.
- [39] Nielsen, I. M.; Janssen, C. L. *Chem. Phys. Lett.* **1999**, *310*, 568 – 576.
- [40] Lee, T. J. *Chem. Phys. Lett.* **2003**, *372*, 362–367.
- [41] Fogueri, U. R.; Kozuch, S.; Karton, A.; Martin, J. M. L. *Theor. Chem. Acc.* **2012**, *132*.
- [42] Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864–B871.
- [43] Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133–A1138.
- [44] Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 1053–1062.
- [45] Perdew, J. P.; Yue, W. *Phys. Rev. B* **1986**, *33*, 8800–8802.
- [46] Reiher, M.; Salomon, O.; Artur Hess, B. *Theor. Chem. Acc.* **2001**, *107*, 48–55.
- [47] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- [48] Grimme, S. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 211–228.

- [49] Johnson, E. R.; Mackie, I. D.; DiLabio, G. A. *J. Phys. Org. Chem.* **2009**, *22*, 1127–1135.
- [50] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
- [51] Warshel, A.; Karplus, M. *J. Am. Chem. Soc.* **1972**, *94*, 5612–5625.
- [52] Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227 – 249.
- [53] Senn, H.; Thiel, W. In *Atomistic Approaches in Modern Biology*; Reiher, M., Ed.; Topics in Current Chemistry, Vol. 268; Springer Berlin Heidelberg, 2007; pp 173–290.
- [54] Senn, H. M.; Thiel, W. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.
- [55] Broyden, C. G. *IMA J. Appl. Math.* **1970**, *6*, 222–231.
- [56] Fletcher, R. *Comput. J.* **1970**, *13*, 317–322.
- [57] Goldfarb, D. *Math. Comp.* **1970**, *24*, 23–26.
- [58] Shanno, D. F. *Math. Comp.* **1970**, *24*, 647–656.
- [59] Li, X.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2*, 835–839.
- [60] Nocedal, J. *Math. Comp.* **1980**, *35*, 773–782.
- [61] Liu, D. C.; Nocedal, J. *Math. Prog.* **1989**, *45*, 503–528.
- [62] Banerjee, A.; Adams, N.; Simons, J.; Shepard, R. *J. Phys. Chem.-US* **1985**, *89*, 52–57.
- [63] Billeter, S. R.; Turner, A. J.; Thiel, W. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177–2186.
- [64] Karplus, M.; McCammon, J. A. *Nat Struct Mol Biol* **2002**, *9*, 646–652.
- [65] Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.
- [66] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- [67] Chang, M. W.; Ayeni, C.; Breuer, S.; Torbett, B. E. *PLoS ONE* **2010**, *5*, e11955.
- [68] Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010**, *31*, 455–461.
- [69] Altun, A.; Kumar, D.; Neese, F.; Thiel, W. *J. Phys. Chem. A* **2008**, *112*, 12904–12910.

- [70] Schöneboom, J. C.; Neese, F.; Thiel, W. *J. Am. Chem. Soc.* **2005**, *127*, 5840–5853.
- [71] Chen, H.; Song, J.; Lai, W.; Wu, W.; Shaik, S. *J. Chem. Theory Comput.* **2010**, *6*, 940–953.
- [72] Geng, C.; Ye, S.; Neese, F. *Angew. Chem.* **2010**, *122*, 5853–5856.
- [73] Lundberg, M.; Borowski, T. *Coord. Chem. Rev.* **2013**, *257*, 277–289.
- [74] McDonald, A. R.; Que, L. *Coord. Chem. Rev.* **2013**, *257*, 414–428.
- [75] Chen, H.; Lai, W.; Shaik, S. *J. Phys. Chem. B* **2011**, *115*, 1727–1742.
- [76] Ye, S.; Geng, C.-Y.; Shaik, S.; Neese, F. *Phys. Chem. Chem. Phys.* **2013**, *15*, 8017–8030.
- [77] Schenk, G.; Pau, M. Y. M.; Solomon, E. I. *J. Am. Chem. Soc.* **2004**, *126*, 505–515.
- [78] TURBOMOLE V6.4 2012, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from <http://www.turbomole.com>.
- [79] Treutler, O.; Ahlrichs, R. *J. Chem. Phys.* **1995**, *102*, 346–354.
- [80] Von Arnim, M.; Ahlrichs, R. *J. Comput. Chem.* **1998**, *19*, 1746–1757.
- [81] Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- [82] Weigend, F. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.
- [83] Sierka, M.; Hogekamp, A.; Ahlrichs, R. *J. Chem. Phys.* **2003**, *118*, 9136–9148.
- [84] Werner, H.-J.; et al.; *MOLPRO, version 2012.1, a package of ab initio programs*; 2012; see <http://www.molpro.net>.
- [85] Knowles, P. J.; Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **1993**, *99*, 5219–5227.
- [86] Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. *WIREs Comput. Mol. Sci.* **2012**, *2*, 242–253.
- [87] Knizia, G.; Adler, T. B.; Werner, H.-J. *J. Chem. Phys.* **2009**, *130*, 054104.
- [88] Allen, F. H. *Acta Crystallogr. Sect. B-Struct. Sci.* **2002**, *58*, 380–388.
- [89] Rohde, J.-U.; In, J.-H.; Lim, M. H.; Brennessel, W. W.; Bukowski, M. R.; Stubna, A.; Münck, E.; Nam, W.; Que, L. *Science* **2003**, *299*, 1037–1039.

- [90] Thibon, A.; England, J.; Martinho, M.; Young, V.; Frisch, J.; Guillot, R.; Girerd, J.-J.; Münck, E.; Que, L.; Banse, F. *Angew. Chem. Int. Ed.* **2008**, *47*, 7064–7067.
- [91] Sinnecker, S.; Svensen, N.; Barr, E. W.; Ye, S.; Bollinger, J. M.; Neese, F.; Krebs, C. *J. Am. Chem. Soc.* **2007**, *129*, 6168–6179.
- [92] Lide, D. R. *CRC handbook of chemistry and physics: a ready-reference book of chemical and physical data*; CRC Press: Boca Raton, Fla., 2009.
- [93] Srnec, M.; Wong, S. D.; England, J.; Que, L.; Solomon, E. I. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 14326–14331.
- [94] Kulik, H. J.; Drennan, C. L. *J. Biol. Chem.* **2013**, *288*, 11233–11241.
- [95] Andrew Jarnuczak; MSci thesis; University of Glasgow; 2011.
- [96] Šali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779–815.
- [97] Fiser, A.; Do, R. K. G.; Šali, A. *Prot. Sci.* **2000**, *9*, 1753–1773.
- [98] Word, J.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. *J. Mol. Biol.* **1999**, *285*, 1735 – 1747.
- [99] Chen, V.; Arendall, W.; Headd, J.; Keedy, D.; Immormino, R.; Kapral, G.; Murray, L.; Richardson, J.; Richardson, D. *Acta Cryst. D* **2010**, *66*, 12–21.
- [100] Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219–10225.
- [101] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- [102] Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- [103] Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- [104] Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129–145.
- [105] Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.
- [106] Sousa da Silva, A. W.; Vranken, W. F. *BMC Research Notes* **2012**, *5*, 367.

- [107] Senn, H. M.; Thiel, S.; Thiel, W. *J. Chem. Theory Comput.* **2005**, *1*, 494–505.
- [108] Schiferl, S. K.; Wallace, D. C. *J. Chem. Phys.* **1985**, *83*, 5203–5209.
- [109] Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J. *PLoS Comput. Biol.* **2012**, *8*, e1002708.
- [110] Blomberg, M. R.; Borowski, T.; Himo, F.; Liao, R.-Z.; Siegbahn, P. E. *Chem. Rev.* **2014**, *114*, 3601–3658.
- [111] de Visser, S. P. *Chem. Commun.* **2007**, 171–173.
- [112] Topol, I. A.; Nemukhin, A. V.; Salnikow, K.; Cachau, R. E.; Abashkin, Y. G.; Kasprzak, K. S.; Burt, S. K. *J. Phys. Chem. A* **2006**, *110*, 4223–4228.
- [113] Diebold, A. R.; Brown-Marshall, C. D.; Neidig, M. L.; Brownlee, J. M.; Moran, G. R.; Solomon, E. I. *J. Am. Chem. Soc.* **2011**, *133*, 18148–18160.
- [114] Hirao, H.; Que, L.; Nam, W.; Shaik, S. *Chem. Eur. J.* **2008**, *14*, 1740–1756.
- [115] Janardanan, D.; Wang, Y.; Schyman, P.; Que, L.; Shaik, S. *Angew. Chem. Int. Ed.* **2010**, *49*, 3342–3345.
- [116] ChemShell, a Computational Chemistry Shell, see www.chemshell.org.
- [117] Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; et al. *J. Mol. Struc.-THEOCHEM* **2003**, *632*, 1–28.
- [118] Smith, W.; Forester, T. *J. Mol. Graphics.* **1996**, *14*, 136 – 141.
- [119] Kästner, J.; Carr, J. M.; Keal, T. W.; Thiel, W.; Wander, A.; Sherwood, P. *J. Phys. Chem. A* **2009**, *113*, 11856–11865.